

GLOSSARY

A

adaptive routing (also, *dynamic adaptive routing, alternate routing, or state-dependent routing*): a routing scheme in which the choice of a route depends on the load level in the network. Adaptive routing makes it possible to utilize network resources more efficiently than with static routing.

Advanced Research Projects Agency (ARPA): a U.S. federal research agency that was responsible for the funding of the network now known as the Internet. The abbreviation DARPA was also used for a while (D stands for Defense).

application: a computer program that performs a significant function directly for a user. Email and IP telephony are typical applications used in the Internet.

Assured Forwarding PHB (AF PHB): a PHB group consisting of four PHB classes. AF PHB group can be used to provide quality differentiation related to various quality aspects.

Asynchronous Transfer Mode (ATM): a connection-oriented, packet switching technology where information is carried in fixed-size cells. The objective of ATM technology has been to provide a ubiquitous network service for every imaginable purpose.

Availability of Quality: the probability that a service can meet a given quality requirement. Availability of Quality is usually measured over a long period of time.

Available Bit Rate (ABR): an ATM service category in which a network continuously informs end systems about the current available capacity in the network. An end system that can adapt its bit rate in accordance with the feedback information is expected to obtain a low cell-loss ratio.

average bit rate: a bit rate of a traffic stream measured over a relatively long period.

B

backbone: the primary connectivity mechanisms of a hierarchical system. The Internet backbone is the part of the network used to connect Internet service providers to each other.

bandwidth: formally, the difference between the highest and lowest frequencies of a transmission channel. In practice, widely used as a general term for the capacity of a transmission channel in bits per second.

best-effort service (BE): a service model in which the network transmits packets without explicit quality guarantees. This is the prevalent Internet service model.

Birds of a Feather (BoF): An informal discussion group to consider a specific issue. In an IETF meeting, BoF is a

session held before the formation of a working group to decide whether the interest is sufficient.

boundary node: a network node that connects one domain to another domain. In a Differentiated Services environment, a boundary node can connect a domain to a customer, to another Differentiated Services domain, or to a domain without Differentiated Services capabilities.

buffer: a temporary storage in which data is held pending an opportunity to complete its transfer. In telecommunication networks, a buffer can be considered a technical implementation of a queue, but often buffer and queue are used as synonyms.

C

call blocking (or *call-blocking probability*): the probability that a call request is rejected because of insufficient network resources. Call-blocking standards and calculations are extensively used for planning telephone networks.

call: an association between two or more end systems connected to a network in order to use network capabilities. In multiple-service networks, a call can consist of a number of connections.

Cell Loss Priority (CLP): a bit in the ATM cell header that indicates two levels of drop priority. A CLP value of 1 in an ATM cell implies that an ATM node can discard the cell during a congestion situation.

cell: the unit of transmission in ATM networks. An ATM cell has a fixed size—a 5-octet header and 48-octet payload.

circuit switching: a communications principle in which a dedicated channel is established between end systems. The channel reserves certain network resources independent of the actual traffic sent through the channel.

Class Selector PHB (CS PHB): a PHB group that provides backward compatibility for systems using the former specification of the TOS field.

Class-based Queuing (CBQ): a queuing method that classifies packets according to certain criteria and reserves a separate queue for each traffic class. The goal of CBQ is to provide appropriated treatment for each traffic class.

classifier: an entity that selects packets according to defined rules. In a Differentiated Services network, a classifier selects the PHB class for a traffic flow.

codepoint (or *DS codepoint [DSCP]*): a specific value of the first six bits in the DS field. The codepoint of a packet informs the network about the PHB of the packet.

congestion: a situation in which there are not enough resources to appropriately handle all tasks. In packet networks, congestion occurs when the load level exceeds either the capacity of a network node or the bandwidth of a link. Congestion may last from milliseconds to weeks.

congestion control: a set of mechanisms used to avoid congestion inside the network. Congestion control includes mechanisms that enable the sender to adjust the transmission rate according to the available bit rate inside the network.

connection: a possibility to transfer information between two or more end systems. Usually, a transmission channel is explicitly established before the actual use of a connection.

Connection Admission Control (CAC): a set of actions during the call setup phase that determines whether a connection request can be accepted or should be rejected. CAC is an indispensable control mechanism in any network that gives quality guarantees.

connection oriented: a communication method in which a connection has to be established before information transfer. ATM is connection-oriented networking technology.

connectionless: a communication method in which information can be sent without first establishing a connection. IP is a connectionless protocol.

Constant Bit Rate (CBR): an ATM service category in which a constant bandwidth is allocated for a traffic stream. CBR can also be used as an attribute for any connection that sends (or is expected to send) traffic with a constant bit rate.

controlled-load service: a service model in which the network permanently offers a quality level similar to that of an unloaded

network. The quality of a controlled load service is expected to be sufficient for most applications except the most demanding real-time applications.

cost efficiency (or *cost effectiveness*): economical in terms of services received for the money spent. Cost efficiency is one factor in this book used to measure whether Differentiated Services technology is providing a profitable business for the service provider and affordable services for customers.

customer service: the relationship between the customer and the service provider, including all issues that significantly affect customer satisfaction.

D

datagram: the basic information unit in the Internet that contains sufficient information to route it from source to destination. In practice, *packet* is often used rather than *datagram* as a general term for a basic information unit.

Differentiated Services (DiffServ): a service paradigm in which quality differentiation is based on the classification and marking of packets rather than explicit resource reservations for individual flows.

domain: a contiguous set of nodes operated under a common service-provisioning principle. A domain is usually managed by a single organization.

drop precedence: a value in a packet header that is used during a congestion

situation to decide which packets are discarded.

DS field: the IPv4 header TOS octet or the IPv6 Traffic Class octet when used in conformance with the Differentiated Services specifications.

dynamic importance: a term used in this book to describe a system that enables customers to request a specific importance level for traffic flows or for individual packets.

Dynamic Real-Time/non-real-time PHB (DRT PHB): a PHB group that consists of two PHB classes, each with six importance levels. A DRT PHB group with appropriate traffic-condition methods can be used to build quality differentiation related to delay, loss probability, and bandwidth.

E

egress node: a collection of functions used to handle outgoing traffic streams from a network domain. Typically, a node is an egress node for some traffic streams and ingress node for some other traffic streams.

equal queuing: a term used in this book for a queuing system that attempts to divide the link capacity evenly among some entities. The entity could be an individual flow or an aggregate traffic stream.

Expedited Forwarding PHB (EF PHB): a PHB group that can be used to build a high-quality, virtual circuit service in a Differentiated Services network.

F

fair queuing: a queuing system that can provide a fair sharing of network resources even when some flows attempt to use as much resources as possible.

fairness: reasonable according to most people's ideas of justice. Fairness is used in this book to emphasize that the fundamental needs of customers are at least as important as technical requirements.

first in, first out (FIFO): a queuing system in which packets are transmitted in the order in which they are received.

flow: an association of packets transmitted between two end systems. In IP networks, packets with the same source address, source port, destination address, and destination port are often considered a flow if the packets are sent within a relatively short period of time.

flow control: a mechanism that enables the sender to adjust the transmission rate according to the available bit at a receiver.

forwarding: an operation performed by a network node in which a packet is received on an input, the right output is determined, and the packet is sent to the right output. There is a clear distinction between routing and forwarding; routing is a supporting function for the forwarding process of packets.

frame: an information unit usually related to the transmission of information over one link rather than through the whole network. *Packet* is often used rather than *frame* as a more general term.

G–H

General Packet Radio Service (GPRS): a specification for data transmission in wireless networks.

guaranteed connections: a term used for a network service used by an individual application with specific quality requirements and duration.

Guaranteed Frame Rate (GFR): an ATM service category in which a specific part of traffic flow is delivered with high probability while the excess traffic obtains best-effort service.

guaranteed service: a service model in which a network attempts to meet specific bandwidth and quality requirements with very high probability. The guaranteed-service model developed by IETF is expected to offer high enough quality for real-time applications.

header: a block of octets in the beginning of a packet that contains control information, such as source and destination addresses.

hop: a path between two network nodes that does not have any significant effect on the characteristics of traffic flows. An ATM VP between two routers is one hop, although the VP goes through several ATM nodes, if those nodes keep the traffic process virtually intact.

I

importance level: information about the relative importance of a packet to be used for traffic-management purposes. In this book, *importance level* is used (rather than *drop precedence* or similar terms) as a general term that does not refer to any specific implementation.

ingress node: a collection of functions used to handle incoming traffic streams to a network domain. Typically, a node is an egress node for some traffic streams and an ingress node for some other traffic streams.

Institute of Electrical and Electronics Engineers (IEEE): a professional society that also makes communication and network standards. IEEE has an integral role of defining standards for local area networks.

Integrated Service Digital Network (ISDN): an evolutionary step of a telephone network to integrate voice and data traffic into the same network. The main application of ISDN related to data traffic is to provide access to the Internet with a maximum bandwidth of 128kbps using ordinary telephone lines.

Integrated Services (IntServ): a service architecture in which a diverse range of services, such as voice, video, and data, are transmitted within the same network infrastructure. In IETF, Integrated Services refers to the effort made by the Integrated Services Working Group to design an advanced service model that may replace best-effort service.

interior node: a network that has links only to nodes within the same domain.

Internet (with capital “I”): a global information network that consists of a large number of smaller internets. The Internet is usually considered to cover networks with public access, whereas large networks with closed access are called intranets.

internet: a set of packet networks interconnected by routers that enable them to function as one unified network.

Internet Draft: a working document submitted by an IETF working group or by individual contributors. An Internet Draft is the first phase of the development of an RFC, but not all Internet Drafts lead to a RFC.

Internet Engineering Task Force (IETF): an organization that provides the coordination of standard and specification development for the Internet.

Internet Protocol (IP): a protocol that provides a connectionless delivery of packets in the Internet.

Internet Protocol version 4 (IPv4): the predominant version of the Internet Protocol today.

Internet Protocol version 6 (IPv6): a new version of the Internet Protocol with several enhanced features, including a larger address space. It is expected that IPv6 will replace IPv4 in the future, although the schedule is still largely uncertain.

Internet service provider (ISP): a company that sells access to the Internet.

interworking: a set of issues related to the capability of different standards, techniques, and mechanisms to work together.

J–L

jitter: a distortion of a signal, such as a flow of packets, in which the original timing relationships are altered. In the context of packet networks, *jitter* is a synonym for *delay variation*.

label: a short, fixed-length identifier in a packet header. The value of a label is usually changed in every node.

label switching: a technique based on labels that is used to alleviate the processing effort of packet forwarding.

leased-line service: a service in which a virtual circuit is leased for exclusive use of a specific customer. In this book, *leased-line service* is also used as a general term to describe a network service with permanent reservation of bandwidth and with high-quality requirements.

link: a physical connection between two network nodes.

local area network (LAN): a data network that spans small distances, up to a few kilometers. LANs are usually administered by a single organization.

M

marking: the process of setting bits in a packet header in order to have an effect on the treatment of the packet. The term *marking* is also used in this book in a

more limited sense: changing the importance level of a packet within a PHB class.

mechanism: a system of parts working together used to achieve a specific result. In Differentiated Services networks, typical parts are queuing disciplines or dropping algorithms, and the result can be a Per-Hop Behavior.

metering: a measuring process of temporal properties of a traffic stream. In this book, metering refers to cases where the result is used for packet marking, shaping, or dropping.

multicast: a technique in which a packet is delivered to more than one destination. Multicast improves the utilization of network resources by minimizing duplication of packets.

multiplexing: a technique for transmitting a number of separate signals over a single channel.

Multiprotocol Label Switching

(MPLS): an IETF working group for developing label-switching standards. The development work of MPLS also includes traffic-management aspects similar to Differentiated Services.

N–O

network dimensioning (or *capacity planning*): a methodology used to manage network resources in a reasonable manner. The timescale of network dimensioning is usually at least a few weeks.

network operator: a company, organization, or person responsible for the operation and management of a network.

network service: the part of customer service that defines the technical characteristics of information transmission through a network.

node (or *network node*): a device attached to a network with capability to make connections to other devices.

Nominal Bit Rate (NBR): a parameter used in DRT PHB and in SIMA that defines the share of network resources obtained by a traffic stream.

non-real-time (NRT): a characteristic of an application or of a service without strict transmission delay and delay-variation constraints.

octet: a group of eight bits. *Byte* is often used as a synonym for *octet*, although in a strict sense byte does not define the number of bits.

operation and management (OAM): a set of functions that provides essential information about the condition of a network and that are used to make operational actions.

P

packet: a generic term used to describe an information unit that contains enough information to transmit it through a network.

packet-loss ratio (P_{loss}): the number of lost packets divided by the number of

total packets. Packet-loss ratio can be defined for different levels of aggregation and for different periods of time.

packet switching: a communication principle in which information is divided into packets that are stored in network nodes before being transmitted forward.

peak rate: a bit rate of a traffic stream measured over a very short period. Peak rate is also used as a traffic parameter that defines the highest rate at which source can send data into the network.

Per-Hop Behavior (PHB): an externally observable forwarding treatment of an aggregate traffic stream in a network node. Conceptually, PHB is located between mechanisms and network services; service providers build network services based on PHBs that are implemented by using mechanisms.

PHB class: a PHB group intended to be applicable for transmitting packets of one application. A PHB class offers basically the same delay characteristics for all flows using the class.

PHB group: a set of one or more PHBs that has to be specified and implemented simultaneously. In this sense, two or more PHB groups can be used in a network domain only if the operator allocates strictly separate network resources for every PHB group.

port: (1) a physical interface to a network node. (2) an identifier used to specify a particular application running on a computer.

predictability of quality: a term used in this book to describe the capability of users to predict the actual quality of network service.

protocol: a formal description of messages and rules to be used by two or more systems to exchange information. Protocols make logical connections between applications on different computers.

Q–R

quality of service (QoS): a set of attributes that can be used to define the network's capability to meet the requirements of users and applications. Packet-loss ratio and delay variation are typical QoS attributes.

Random Early Detection (RED): a queuing principle in which packets are dropped randomly before the queue becomes full. The main goal of RED is to improve the efficiency of TCP congestion control.

real-time (RT): characteristic of an application or a service with strict transmission delay and delay-variation constraints.

Request For Comment (RFC): a document produced by IETF that contains information about the Internet. Internet standards are written in the form of an RFC, but only some RFCs are actually standards.

Resource Reservation Protocol (RSVP): a protocol used in the Internet to reserve resources for flows with specific QoS requirements.

resource sharing: a network service model in which resources are divided among all active users according to specific rules. The actual amount of bandwidth available for a user depends on the overall level of demand.

robustness: the capability to work properly under harsh conditions. Robustness is used in this book to emphasize the need to build a system that works appropriately in diverse conditions, even when some users are malicious.

router: a network node that performs routing functions and forwards packets to other nodes based on the routing information.

routing: a process in network nodes that exchanges and maintains information about paths to various destinations. The results of routing are used to forward packets through the network.

S

scheduling discipline: an algorithm that determines the order in which service requests are served.

service-level agreement (SLA): the formal part of the relationship between a service provider and a customer. The SLA contains assurances related to various issues such as service availability and transmission delays.

shaping: a mechanism that delays packets of a traffic stream to achieve better network efficiency or to ensure conformance with a traffic profile.

signaling: a process that allows end systems and network nodes to exchange information. Particularly, signaling is used for establishing and releasing connections.

Simple Integrated Media Access (SIMA): a Differentiated Services model that includes specifications for network mechanisms, traffic conditioning, and pricing.

starvation: a situation in which a traffic stream with low priority does not get any service because other traffic streams with higher priority use all the resources.

statistical multiplexing: a multiplexing principle in which there is a possibility that network resources are not sufficient for transmitting all signals. The objective of statistical multiplexing is to improve the utilization of network resources.

Stochastic Fairness Queuing (SFQ): a variant of fair queuing in which a special hash function is used to map a flow to one of a set of queues. The number of required queues is smaller than in a WFQ system.

switch: a network node that makes the forwarding decision based on a label rather than on the full destination address. Sometimes *switch* is used as a general term for all devices that transfer information from input to output.

T

throughput: the transmission capacity used by traffic streams in a channel in bits per second. *Bandwidth* is often used as a synonym for *throughput*.

token bucket: a control algorithm that regulates the bandwidth used by a traffic stream but allows some level of burstiness. *Leaky bucket* is often (but not always) used as a synonym for *token bucket*.

traffic conditioning: a set of mechanisms used to control and modify traffic flows. Traffic conditioning includes metering, marking, shaping, and dropping.

traffic handling: a set of mechanisms used by a network operator to take care of traffic in a network.

traffic management: a set of principles and mechanisms that allows the network operator to efficiently utilize network resources and to meet customer requirements.

traffic profile: a description of the temporal properties of a traffic stream. Traffic profile includes parameters (such as peak rate and maximum burst size) that are used for traffic-conditioning purposes.

traffic stream: a set of one or more flows with some common characteristics. Flows can be classified in to a stream, for instance, based on the customer group or on the quality requirements.

Transmission Control Protocol (TCP): a protocol that provides reliable connections in the Internet. The majority of traffic in the present Internet uses TCP.

Type of Service (TOS): a field in the IPv4 packet header designed to indicate the preferred treatment of the packet. The former specification of TOS bits has been

replaced by Differentiated Services specifications.

U–Z

Unspecified Bit Rate (UBR): an ATM service category without QoS or bandwidth guarantees. UBR service is similar to the best-effort service used in the Internet.

User Datagram Protocol (UDP): a simple connectionless protocol used to transport packets in the Internet. As opposed to TCP, UDP does not provide any congestion-control mechanisms itself.

Variable Bit Rate (VBR): an ATM service category intended for applications with bandwidth requirements that vary with time. There are two subcategories: rt-VBR is intended for real-time applications and nrt-VBR is intended for non-real-time applications.

versatility: the capability to do many things competently. The term *versatility* is used in this book to emphasize the need of a single consistent system to realize various services and characteristics.

virtual circuit (VC): a technique that provides connection-oriented service regardless of the underlying network structure.

virtual path (VP): a set of virtual circuits established for traffic-management purposes.

virtual private network (VPN): a private network in which network nodes are at

least partially connected through a public network. VPNs use secure protocols to ensure that unauthorized parties do not intercept data transfers.

Weighted Fair Queuing (WFQ): a variant of fair queuing in which each traffic flow can be given a specific proportion of the network capacity.

Weighted Random Early Detection (WRED): a variant of RED in which the packet-discarding probability depends on the characteristics of the flows or on the importance level of the packet.

World Wide Web (WWW or Web): a distributed information system used in the Internet.

X.25: a connection-oriented packet-switching protocol.

