International Telecommunication Union

ITU-T Y.1221

TELECOMMUNICATION
STANDARDIZATION SECTOR
OF ITU

(06/2010)

SERIES Y: GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

Internet protocol aspects – Architecture, access, network capabilities and resource management

Traffic control and congestion control in IP-based networks

Recommendation ITU-T Y.1221

ITU-T Y-SERIES RECOMMENDATIONS

GLOBAL INFORMATION INFRASTRUCTURE, INTERNET PROTOCOL ASPECTS AND NEXT-GENERATION NETWORKS

GLOBAL INFORMATION INFRASTRUCTURE	
General	Y.100–Y.199
Services, applications and middleware	Y.200–Y.299
Network aspects	Y.300–Y.399
Interfaces and protocols	Y.400–Y.499
Numbering, addressing and naming	Y.500–Y.599
Operation, administration and maintenance	Y.600–Y.699
Security	Y.700–Y.799
Performances	Y.800–Y.899
INTERNET PROTOCOL ASPECTS	
General	Y.1000–Y.1099
Services and applications	Y.1100–Y.1199
Architecture, access, network capabilities and resource management	**Y.1200–Y.1299**
Transport	Y.1300–Y.1399
Interworking	Y.1400–Y.1499
Quality of service and network performance	Y.1500–Y.1599
Signalling	Y.1600–Y.1699
Operation, administration and maintenance	Y.1700–Y.1799
Charging	Y.1800–Y.1899
IPTV over NGN	Y.1900–Y.1999
NEXT GENERATION NETWORKS	
Frameworks and functional architecture models	Y.2000–Y.2099
Quality of Service and performance	Y.2100–Y.2199
Service aspects: Service capabilities and service architecture	Y.2200–Y.2249
Service aspects: Interoperability of services and networks in NGN	Y.2250–Y.2299
Numbering, naming and addressing	Y.2300–Y.2399
Network management	Y.2400–Y.2499
Network control architectures and protocols	Y.2500–Y.2599
Future networks	Y.2600–Y.2699
Security	Y.2700–Y.2799
Generalized mobility	Y.2800–Y.2899
Carrier grade open environment	Y.2900–Y.2999

For further details, please refer to the list of ITU-T Recommendations.

Recommendation ITU-T Y.1221

Traffic control and congestion control in IP-based networks

Summary

Recommendation ITU-T Y.1221 provides a general description as well as objectives and procedures for traffic control and congestion control for IP-based networks. In particular, it describes the concepts of the traffic contract between a user and the network. It specifies the IP transfer capabilities (IPTCs) including, for each IPTC, the service model, the associated traffic patterns and conformance definition.

History

Edition	Recommendation	Approval	Study Group
1.0	ITU-T Y.1221	2002-03-16	13
1.1	ITU-T Y.1221 (2002) Amend. 1	2004-03-29	13
1.2	ITU-T Y.1221 (2002) Amend. 2	2005-11-29	12
1.3	ITU-T Y.1221 (2002) Amend. 3	2007-10-11	12
2.0	ITU-T Y.1221	2010-06-29	12

FOREWORD

The International Telecommunication Union (ITU) is the United Nations specialized agency in the field of telecommunications, information and communication technologies (ICTs). The ITU Telecommunication Standardization Sector (ITU-T) is a permanent organ of ITU. ITU-T is responsible for studying technical, operating and tariff questions and issuing Recommendations on them with a view to standardizing telecommunications on a worldwide basis.

The World Telecommunication Standardization Assembly (WTSA), which meets every four years, establishes the topics for study by the ITU-T study groups which, in turn, produce Recommendations on these topics.

The approval of ITU-T Recommendations is covered by the procedure laid down in WTSA Resolution 1.

In some areas of information technology which fall within ITU-T's purview, the necessary standards are prepared on a collaborative basis with ISO and IEC.

NOTE

In this Recommendation, the expression "Administration" is used for conciseness to indicate both a telecommunication administration and a recognized operating agency.

Compliance with this Recommendation is voluntary. However, the Recommendation may contain certain mandatory provisions (to ensure e.g., interoperability or applicability) and compliance with the Recommendation is achieved when all of these mandatory provisions are met. The words "shall" or some other obligatory language such as "must" and the negative equivalents are used to express requirements. The use of such words does not suggest that compliance with the Recommendation is required of any party.

INTELLECTUAL PROPERTY RIGHTS

ITU draws attention to the possibility that the practice or implementation of this Recommendation may involve the use of a claimed Intellectual Property Right. ITU takes no position concerning the evidence, validity or applicability of claimed Intellectual Property Rights, whether asserted by ITU members or others outside of the Recommendation development process.

As of the date of approval of this Recommendation, ITU had received notice of intellectual property, protected by patents, which may be required to implement this Recommendation. However, implementers are cautioned that this may not represent the latest information and are therefore strongly urged to consult the TSB patent database at http://www.itu.int/ITU-T/ipr/.

CONTENTS

Recommendation ITU-T Y.1221

Traffic control and congestion control in IP-based networks

1 Scope

This Recommendation describes traffic control and congestion control procedures for IP-based networks. Such control procedures are necessary in order to support services with quality of service (QoS), where the QoS is negotiated between a user and the network.

Traffic control refers to all network actions aiming to meet the negotiated performance objectives in an IP-based network and to allow the avoidance of congested conditions.

Congestion control refers to all network actions to minimize the intensity, spread and duration of congestion.

This Recommendation provides a general description, as well as objectives and procedures, for traffic control and congestion control. In particular, it describes the concepts of the traffic contract between a user and the network. It specifies the IP transfer capabilities (IPTCs) including, for each IPTC, the service model, the associated traffic patterns and conformance definition.

2 References

The following ITU-T Recommendations and other references contain provisions which, through reference in this text, constitute provisions of this Recommendation. At the time of publication, the editions indicated were valid. All Recommendations and other references are subject to revision; users of this Recommendation are therefore encouraged to investigate the possibility of applying the most recent edition of the Recommendations and other references listed below. A list of the currently valid ITU-T Recommendations is regularly published. The reference to a document within this Recommendation does not give it, as a stand-alone document, the status of a Recommendation.

[ITU-T I.371] Recommendation ITU-T I.371 (2000), *Traffic control and congestion control in B-ISDN*.

[ITU-T Y.1241] Recommendation ITU-T Y.1241 (2001), *Support of IP-based services using IP transfer capabilities*.

[ITU-T Y.1540] Recommendation ITU-T Y.1540 (1999), *Internet protocol data communication service – IP packet transfer and availability performance parameters*.

[ITU-T Y.1541] Recommendation ITU-T Y.1541 (2002), *Network performance objectives for IP-based services*.

3 Definitions

3.1 Terms defined elsewhere

None.

3.2 Terms defined in this Recommendation

This Recommendation defines the following terms:

3.2.1 conformance: Conformance is the application of one or more criteria, at a given standardized interface, to a flow.

3.2.2 congestion: Congestion is defined as a state of network elements (e.g., router, switches) in which the network is not able to meet the network performance objectives and the negotiated QoS commitments for the already established flows.

3.2.3 IP congestion control: Refers to all network actions to minimize the intensity, spread and duration of congestion.

3.2.4 IP flow: An IP flow at a given interface is defined as the occurrence at that interface of the set of IP packets which match a given classification (see clause 7.1.2).

3.2.5 IP traffic control: Refers to network actions aiming to meet the network performance objectives and negotiated QoS commitments.

3.2.6 IP transfer capability: An IP transfer capability is a set of network capabilities provided by an IP-based network to transfer IP flows.

3.2.7 overload: Overload is defined as a state of a network element in which buffer overflow results in packet discard for flows with no QoS commitments.

3.2.8 packet classification: The process of distinguishing IP packets for the purpose of applying appropriate traffic control and congestion control mechanisms (see clause 7.1.1).

3.2.9 traffic contract: For a given IP flow, the selected IP transfer capability (see clause 8), the traffic descriptor at a given interface and the QoS class [ITU-T Y.1541] define the traffic contract at that interface (see clause 7.1.5).

3.2.10 traffic descriptor: A traffic descriptor is the set of traffic parameters that is used to capture the traffic characteristics of an IP flow at a given standardized interface as part of the traffic contract (see clause 7.1.4).

3.2.11 traffic parameter: A traffic parameter describes one aspect of a flow (see clause 7.1.3).

4 Abbreviations and acronyms

This Recommendation uses the following abbreviations and acronyms:

BE	Best Effort
Bp	Bucket size of peak token bucket
Bs	Bucket size of sustainable token bucket
CBR	Constant Bit Rate
CDBW	Conditionally Dedicated Bandwidth (IP transfer capability)
CDV	Cell Delay Variation
CSPF	Constrained Shortest Path First
DBW	Dedicated Bandwidth (IP transfer capability)
DS	Differentiated Services (IP header field)
DSBW	Delay-sensitive Statistical Bandwidth (IP transfer capability)
FRR	Fast Reroute
GBRA	Generic Byte Rate Algorithm
GCRA	Generic Cell Rate Algorithm
IP	Internet Protocol
IPDV	IP Delay Variation
IPLR	IP Loss Ratio

IPTC IP Transfer Capability

IPTD IP Transfer Delay

LR Line Rate

LSP Label Switched Path

MPLS Multiprotocol Label Switching

OSPF Open Shortest Path First

PC Parameter Control

PCR Peak Cell Rate

QoS Quality of Service

Rp Rate of peak token bucket

Rs Rate of sustainable token bucket

SBW Statistical Bandwidth (IP transfer capability)

SDP Session Description Protocol

SIP Session Initiation Protocol

TAT Theoretical Arrival Time

TB Token Bucket

TC Transfer Capability

ToS Type of Service (IP header field, see DS)

5 Conventions

This Recommendation uses the following conventions:

- *B*: Bucket size of a token bucket;
- *M*: Maximum allowed packet size;
- *N*: Size (in bytes) of an IP packet;
- *R*: Rate of a token bucket.

6 General

The primary role of traffic control and congestion control procedures is to protect the IP network and the traffic entering the network in order to achieve the IP network performance objectives and QoS commitments. Traffic and congestion control allows the use of IP network resources to be optimized.

In IP-based networks, congestion is defined as a state of network elements (e.g., routers, switches) in which the network is not able to meet the network performance objectives and the negotiated QoS commitments for the established flow. Congestion is to be distinguished from the state where buffer overflow causes IP packet loss, but the negotiated quality of service is still met. For services with no QoS commitments, such as best effort services, the term overload is used rather than congestion.

This Recommendation defines a set of traffic control and congestion control capabilities. It may be appropriate to consider additional sets of such capabilities, for which additional traffic control mechanisms will be used to achieve increased network efficiency.

It should be noted that traffic control procedures apply to IP packet flows. For some IP flows, the network commits to meet QoS objectives, assuming the user-generated IP packet flow conforms to a traffic contract. For IP flows without a traffic contract between the user and the network, the network may provide a best effort service.

7 Traffic parameters and descriptors

7.1 Definitions

7.1.1 IP packet classification

Typically, IP packets from many different sources may cross a given interface to many different destinations. For the purpose of IP traffic control and congestion control, it is essential that not all IP packets are treated the same way but differently, depending on the objectives and on the commitments made. Therefore, it is useful to distinguish IP packets into relevant categories. The process of distinguishing IP packets for the purpose of applying appropriate traffic control and congestion control mechanisms is called IP packet classification. IP packet classification is based on information in the following IP header fields: source address, destination address [b-IETF RFC 791] and TOS/DS field [b-IETF RFC 791]/[b-IETF RFC 2474]. The level of detail in the classification may differ, depending on its intended use. The following examples illustrate this:

– Example 1 – In order to recognize IP packets which flow from a given source address to a given destination address, the entire length of the source and destination address fields is used in the classification.

– Example 2 – In order to recognize IP packets which flow from a given source address to a given destination address with a given service level, the entire length of the source and destination address fields and the DS field is used in the classification.

– Example 3 – In order to recognize IP packets which flow from a source in one given subnet to a destination in another given subnet, the respective subnet masks are applied to the source and destination address fields.

NOTE – In this version of this Recommendation, only the packet classification for IP version 4 has been addressed. Classification for IPv6 and MPLS is for further study.

7.1.2 IP flow and IP subflow

An IP flow at a given interface is defined as the occurrence at that interface of the set of IP packets that match a given classification (see clause 7.1.1). An IP flow may consist of packets from a single application session, or it may be an aggregation comprising the combined traffic from a number of application sessions. When a classification may be subdivided into different sub-classifications (separate or overlapping), different IP subflows may be recognized in the corresponding IP flow.

7.1.3 Traffic parameter

A traffic parameter describes one aspect of a flow. It may be qualitative or quantitative. A traffic parameter may for example describe the peak bit rate, the peak packet rate, the average bit rate, the average or maximum packet size, the (average or maximum) burst length of a flow, etc.

7.1.4 Traffic descriptor

A traffic descriptor is the set of traffic parameters that is used to capture the traffic characteristics of an IP flow at a given standardized interface as part of the traffic contract (see clause 7.1.5).

7.1.5 Traffic contract

For a given IP flow, the selected IP transfer capability (see clause 8), the traffic descriptor at a given interface (see clause 7.1.4) and the QoS class (see [ITU-T Y.1541]) define the traffic contract at that interface.

7.2 Requirements on traffic parameters and traffic descriptors

Any traffic parameter to be involved in a traffic descriptor should:

– have the same interpretation on both sides of an interface;

– be meaningful in resource allocation schemes to meet network performance requirements;

– be enforceable by the parameter control (PC), see clause 9.1.3.

7.3 Traffic parameter specifications

7.3.1 Reference configuration

For further study.

7.3.2 Traffic parameter description

The following traffic parameters are defined, which may be used in traffic descriptors.

7.3.2.1 Maximum allowed packet size

The maximum allowed packet size, *M*, is expressed in bytes. It is a mandatory traffic parameter for each IP transfer capability (IPTC).

7.3.2.2 GBRA or token bucket

The generic byte rate algorithm (GBRA) or token bucket (TB), as described in Annex A, is used to characterize a rate and the associated burstiness. Both concepts are equivalent and use the following set of two parameters:

– the rate, *R*, expressed in bytes/s;

– the bucket size, *B*, expressed in bytes.

The traffic descriptor may contain zero or more token buckets (with the respective values of *R* and *B*).

The current set of IP transfer capabilities (see clause 8) recognizes a peak token bucket (TB) (with parameters peak rate, *Rp*, and peak bucket size, *Bp*) and a sustainable TB (with parameters sustainable rate, *Rs*, and sustainable bucket size, *Bs*).

8 IP transfer capabilities

An IP transfer capability is a set of network capabilities provided by IP-based networks to transfer IP packets. For each IP transfer capability, the service model, traffic descriptor, conformance definition and any QoS commitments are defined. An IP transfer capability is supported by a set of traffic control and congestion control functions.

In order to offer multiple classes of QoS to multiple applications and to optimize the usage of network resources, IP-based networks should be capable of providing multiple transfer capabilities.

Five IP transfer capabilities are defined:

– dedicated bandwidth (DBW) IP transfer capability;

– statistical bandwidth (SBW) IP transfer capability;

– best-effort (BE) IP transfer capability;

– delay-sensitive statistical bandwidth (DSBW) transfer capability;

– conditionally dedicated bandwidth (CDBW) transfer capability.

This set of IP transfer capabilities are based on current IP service models and it may be extended in the future.

8.1 Dedicated bandwidth transfer capability

8.1.1 Description

The dedicated bandwidth (DBW) transfer capability is intended to support applications with stringent delay requirements. It aims to support the guaranteed and timely delivery of IP packets along the end-to-end path of the network.

The DBW transfer capability strives for compatibility with the guaranteed service [b-IETF RFC 2212] and the end-to-end services based on the expedited forwarding per-hop behaviour [b-IETF RFC 2598].

8.1.2 Service model

The DBW IPTC can be used by applications that characterize the traffic with a single token bucket.

The commitment made by the network is that the negotiated IP QoS is assured to all IP packets when all packets are conforming to the conformance tests. The DBW user should expect that (possibly all) non-conforming packets be discarded by the network.

The DBW capability can be associated with specified loss commitments (IP loss ratio (IPLR)) and specified delay commitments (IP transfer delay (IPTD) and IP delay variation (IPDV)), (see [ITU-T Y.1541]).

The network does not fragment packets. In addition, the network commits to attempt, as long as possible (e.g., until there is a need to reroute the flow), to maintain packet sequence integrity.

8.1.3 Traffic descriptor

The traffic descriptor consists of the following parameters:

– The peak rate, *Rp*, and peak bucket size, *Bp*, as specified in clause 7.3.2.2.

– The maximum allowed packet size, *M*, as specified in clause 7.3.2.1.

8.1.4 Conformance definition

An IP packet is conforming if the arrival conforms to the following two parts:

– the arrival is conforming to the GBRA(*Rp,Bp*);

– the actual packet length does not exceed the maximum allowed packet size, *M*.

The GBRA is updated for conforming packets only.

8.1.5 QoS commitments

The DBW capability may be associated with specified loss commitments and specified delay commitments.

If all packets are conforming, the QoS commitments apply to all IP packets. The DBW user should expect that (possibly all) non-conforming packets be discarded by the network. If not all packets are conforming, the network may choose to commit QoS to some of the packets, for example to a volume of packets that is conforming.

8.2 Statistical bandwidth transfer capability

8.2.1 Description

The statistical bandwidth (SBW) transfer capability is intended to support applications that do not have stringent delay requirements. It aims to support the guaranteed delivery of IP packets along the end-to-end path of the network.

The SBW transfer capability strives for compatibility with the controlled-load network element service [b-IETF RFC 2211] and the end-to-end services based on the assured forwarding per-hop behaviour [b-IETF RFC 2597].

8.2.2 Service model

The SBW transfer capability provides a specified sustainable rate (*Rs*) for non-real time applications with limited burst duration, with the expectation that traffic in excess of GBRA(*Rs*,*Bs*) will be delivered within the limits of available resources.

The following two examples describe the commitment the SBW user will receive:

– If the user sends conforming packets at a constant rate that is less than or equal to the *Rs*, then the commitments is that all these packets are delivered across the network, corresponding to the associated QoS class.

– If the user has not sent packets for a long time and the user sends conforming packets in a burst with a duration that does not exceed the bound set by GBRA(*Rs*,*Bs*), then the commitments is that all these packets are delivered across the network, corresponding to the associated QoS class.

The SBW capability also allows the user to send conforming packets in excess of the GBRA(*Rs*,*Bs*), but traffic that exceeds this bound will only be delivered within the limits of available resources.

The SBW capability may be associated with a specified packet loss commitment.

The network does not fragment packets. In addition, the network commits to attempt, as long as possible (e.g., until there is a need to reroute the flow), to maintain packet sequence integrity.

8.2.3 Traffic descriptor

The traffic descriptor consists in:

– the peak rate, *Rp*, and the peak bucket size, *Bp*, as specified in clause 7.3.2.2;

– the sustainable rate, *Rs*, and the sustainable token bucket size, *Bs*, as specified in clause 7.3.2.2;

– the maximum allowed packet size, *M*, as specified in clause 7.3.2.1.

8.2.4 Conformance definition

An IP packet is conforming if the arrival conforms to the following three parts:

– the arrival is conforming to the peak GBRA(*Rp*,*Bp*);

– the arrival is conforming to the sustainable GBRA(*Rs*,*Bs*);

– the actual packet length does not exceed the maximum packet size, *M*.

The GBRAs are updated in coordinated mode (see Annex B) for conforming packets only.

8.2.5 QoS commitments

The SBW capability may be associated with specified loss commitments.

If all packets are conforming, the QoS commitments apply to all packets. Otherwise, the QoS commitments apply to a number of bytes in conforming packets. Non-conforming traffic will be delivered within the limits of available resources.

8.3 Best effort transfer capability

8.3.1 Description

The best effort (BE) IP transfer capability is intended to support applications which do not have stringent loss or delay requirements.

8.3.2 Service model

The service model for the best effort IPTC requires that available resources be used for forwarding packets of best effort flows. Even though there are no QoS commitments specified, the expectation is that packets be delivered provided that sufficient resources are available.

8.3.3 Traffic descriptor

The maximum allowed packet size, *M*.

8.3.4 Conformance definition

Conformance to the maximum packet size *M* is required.

8.3.5 QoS commitments

There is no absolute QoS requirement for this transfer capability.

8.4 Delay-sensitive statistical bandwidth transfer capability

8.4.1 Description

The delay-sensitive statistical bandwidth (DSBW) transfer capability is intended to support applications that do not have requirements on delay variation. It aims to support the guaranteed and timely delivery of IP packets across the end-to-end path of the network.

The DSBW transfer capability strives for compatibility with the controlled-load network element service [b-IETF RFC 2211] and the end-to-end services based on the assured forwarding per-hop behaviour [b-IETF RFC 2597].

8.4.2 Service model

The DSBW transfer capability provides a specified sustainable rate (*Rs*) for non-real time applications with limited burst duration, with the expectation that traffic in excess of GBRA(*Rs*,*Bs*) will be delivered within the limits of available resources.

The following two examples describe the commitment the DSBW user will receive:

– If the user sends conforming packets at a constant rate that is less than or equal to *Rs*, then the commitment is that packets are delivered in a timely way across the network, with performance corresponding to the associated QoS class (see Table VI.1 of [ITU-T Y.1541]).

– If the user has not sent packets for a long time, and the user sends conforming packets in a burst with a duration that does not exceed the bound set by GBRA(*Rs*,*Bs*), then the commitment is that packets are delivered in a timely way across the network, with performance corresponding to the associated QoS class.

The DSBW capability also allows the user to send conforming packets in excess of the GBRA(*Rs*,*Bs*), but traffic that exceeds this bound will only be delivered within the limits of available resources.

The DSBW capability may be associated with a specified packet loss commitment and packet transfer delay commitment to assure timely packet delivery.

The network does not fragment packets. In addition, the network commits to attempt, as long as possible (e.g., until there is a need to reroute the flow), to maintain packet sequence integrity.

8.4.3 Traffic descriptor

The traffic descriptor consists of:

– the peak rate *Rp* and the peak bucket size, *Bp*, as specified in clause 7.3.2.2;

– the sustainable rate, *Rs*, and the sustainable token bucket size, *Bs*, as specified in clause 7.3.2.2;

– the maximum allowed packet size, *M*, as specified in clause 7.3.2.1.

8.4.4 Conformance definition

An IP packet is conforming if the arrival conforms to the following three parts:

– the arrival is conforming to the peak GBRA(*Rp*,*Bp*);

– the arrival is conforming to the sustainable GBRA(*Rs*,*Bs*);

– the actual packet length does not exceed the maximum packet size, *M*.

The GBRAs are updated in coordinated mode (see Annex B) for conforming packets only.

8.4.5 QoS commitments

The SDBW capability may be associated with specified loss commitments.

If all packets are conforming, the QoS commitments apply to all packets. Otherwise, the QoS commitments apply to a number of bytes in conforming packets. Non-conforming traffic may be delivered within the limits of available resources, or they may be discarded, at the discretion of the network provider.

Table 1 summarizes the mapping between the QoS classes of [ITU-T Y.1541] and the transfer capabilities of this Recommendation if this proposal is followed.

Table 1 – Summary mapping table from ITU-T Y.1541 QoS classes to ITU-T Y.1221 transfer capabilities

	QoS classes from [ITU-T Y.1541]					
	Class 0	**Class 1**	**Class 2**	**Class 3**	**Class 4**	**Class 5**
Transfer capabilities from ITU-T Y.1221	DBW		DSBW			BE

8.5 Conditionally dedicated bandwidth transfer capability

8.5.1 Description

The conditionally dedicated bandwidth (CDBW) transfer capability is intended to support applications with stringent delay and varying (application-dependent) loss requirements. However, it is aimed at applications that usually require very low losses or which can tolerate infrequent and short-duration bursts of losses. It would not be likely that these applications would include broadcast quality video, but it could support voice and lower quality video (including, potentially, home movie) transmissions.

CDBW aims to support, with a minimum of negotiation using simplified in-band signalling:

– on most flows, throughout their duration, very low packet losses; roughly equivalent to the network "acceptance" of such flows into this QoS commitment category;

– a worse level of packet loss on (typically) a small percentage of flows that cannot yet be accepted into the above category, but are not denied a transmission path;

– low delay of IP packets along an edge-to-edge path of the network.

The CDBW transfer capability recognizes the existence of one or several IP flows sharing a common physical link. An IP flow consists of a sequence of one or more packets that have the same source IP address, destination IP address, source and destination port numbers and experimental/ diffserv value. Every IP flow of CDBW has a policed maximum rate. Flows exceeding this maximum rate should have their excess packets discarded.

Each flow is assigned a flow state at each network node appropriate to how it will be handled. There are only two possible states: "discard first" and "discard last". Flows in the "discard last" state always have preemptive priority with respect to the available link capacity. In other words, they may rob flows in the "discard first" state if necessary, but subject to the policed maximum rate.

The current flow state of a flow is not conveyed from node-to-node or from network-to-user. Users of this service will expect the provider to make reasonable efforts to ensure that they are infrequently and preferably randomly selected to be in the "discard first" state at the time of network congestion. Furthermore, users will expect that instances of network congestion are infrequent due to appropriate capacity management.

Flows in the "discard first" state are typically, but not necessarily, the latest flows to start. A flow that starts and is assigned a "discard first" state may progress to the "discard last" state subject to the policy conditions that apply. Different rules may be used by administrations governing how and when a flow may progress to the "discard last" state. Some rules may be captured in this Recommendation for information only. Similarly, not all flows that are the latest to start need be assigned to the "discard first" state. Certain flows may be treated preferentially, as determined by the service options of the provider.

Signalling aspects of the CDBW transfer capability will not be described in detail in this Recommendation. In summary there is:

– an in-band signal termed the "Start Packet", recognized by network nodes supporting the CDBW transfer capability that indicates that a new flow is starting and specifies its requested maximum rate and any preference-level request. Authentication information is also included;
– an in-band acknowledgement packet, that confirms that all nodes along the edge-to-edge route have either accepted, rejected, or modified the request;
– an in-band congestion notification packet sent to the receiving edge in the event that packet discards are applied to a flow.

A receiving user who does not at any time receive any congestion notification message relating to a specific flow, and yet perceives packet losses on that flow above the loss limits advertised for this service, would be entitled to complain.

The aim of this transfer capability is not to spread losses across all flows during moments of congestion but, instead, to focus losses consistently on a small set of flows and provide congestion notification signals to the receiving end of such flows.

If, during moments of heavy congestion, all packets from flows in the "discard first" state are being discarded, and if this is insufficient to recover from congestion, then different implementations may limit the extent of packet losses spread among the "discard last" group. For example (and again for information), an implementation may randomly select a small number of "discard last" flows and change their state to "discard first".

The percentage of flows which, at any one time, are in the "discard first" state is a network provider's choice. A qualitative aim for this service is that any one end-user should not experience this bandwidth loss frequently. Flows that request, and are accepted to have, a high preference marking almost never experience discards.

Figure 1 shows how the parameter "% of discard first flows" shifts the service commitment from best effort towards low loss commitments achieved 100% of the time, as experienced by the user.

It can be seen from this figure that CDBW focuses on keeping as many flows as possible free of packet losses for their duration. To achieve this experience purely using best effort principles would require significant levels of over-provision. On the other hand, "circuit mode" principles create the possible experience of an unacceptable call rejection rate unless sufficient over-provision is used to accommodate the busy hour calling rate. CDBW operates with roughly the same level of over-provision as "Circuit Mode" (thus keeping the number of flows that receive congestion notifications low), but does not reject the flows. It takes advantage of any variable bit-rate aspects of these flows that imply that the capacity demanded is less than the sum of the flow peak rates. Furthermore, the transmission path remains available for flows that have received a congestion notification on the grounds that an application would rather have this than outright rejection.

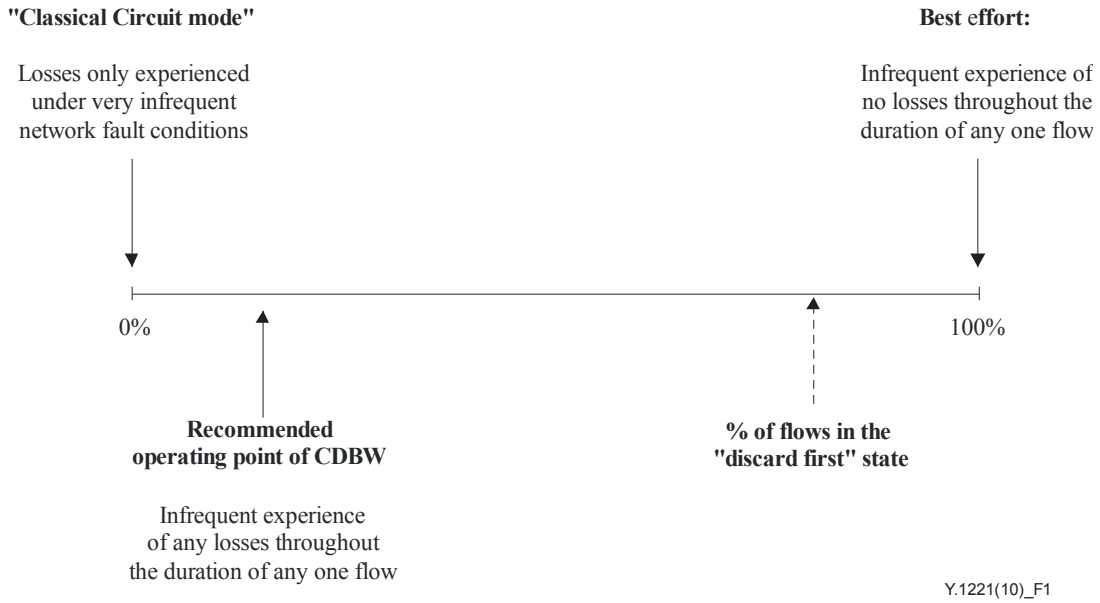


Figure 1 – CDBW transition from a best effort service to imperceptibly near a "circuit mode" service

8.5.2 Service model

The commitment made by the network is conditional depending on the designated flow state. There is a loose analogy between these flow states and "accepted/rejected" calls but, instead of rejection being outright and immediate, a flow may continue in the "discard first" state and the associated application may choose to compensate for any bursts of losses as and when they occur. These may frequently turn out to be only of short duration (or may not occur at all – there is no certainty that flows in the "discard first" state will experience any packet losses).

Therefore, very low packet loss levels are assured for all flows assigned to the preemptive priority "discard last" state. The network will signal a "discard first" receiving user only when its flow is about to experience packet discards. For such users (and associated applications), this signal indicates that, while the flow has not been totally rejected, the very low loss guarantee part of the negotiated QoS cannot be supported. It is up to the application whether to choose to terminate the flow or continue (possibly with a change of coding).

The CDBW user should expect that all non-conforming packets will be discarded by the network.

The CDBW capability for "discard last" flows can be associated with specified loss commitments (IP loss ratio, IPLR) and specified delay commitments (IP transfer delay, IPTD and IP delay variation, IPDV) (see [ITU-T Y.1541]).

The network does not fragment packets. In addition, the network commits to attempt, as long as possible (e.g., until there is a need to reroute the flow), to maintain packet sequence integrity.

8.5.3 Traffic descriptor

The peak rate of the flow is the only mandatory element of the traffic descriptor.

A traffic descriptor should be associated with each flow of this service type coming across a UNI or NNI. This traffic descriptor consists of the following parameters:

– The peak rate, *Rp*, and peak bucket size, *Bp*.

– The maximum allowed packet size, *M*.

8.5.4 Conformance definition

An IP packet is conforming if the aggregate arrival conforms to the following two parts:

– the arrival is conforming to the GBRA(*Rp*,*Bp*);

– the actual packet length does not exceed the maximum allowed packet size, *M*.

The GBRA is updated for conforming packets only.

8.5.5 QoS commitments

The CDBW transfer capability may be associated with specified loss commitments and specified delay commitments, thus making it suitable for the provision of QoS Class 0 and Class 1, within the limits set by the conditional nature of this transfer capability.

If all packets are conforming, the QoS commitments apply to all IP packets of all flows accepted by the network as "discard last". For flows only accepted by the network to be allowed to continue as "discard first" flows, the loss guarantee part of the negotiated QoS does not apply.

9 Functions for traffic control, congestion control, and overload treatment

Traffic control refers to a set of functions that control the flow of packets via a series of functions such as admission control, network resource management and traffic parameter control. The main objective of traffic control is to satisfy user requirements such as quality of service while still supporting efficient network utilization.

As opposed to traffic engineering, traffic control is accomplished in a short time-scale. Therefore, a well-established and automated mechanism is to be provided to control the flow of traffic into the network and out of the network.

Under normal operation, i.e., when no network failures occur, functions referred to as traffic control functions in this Recommendation are intended to avoid network congestion.

However, congestion may occur, e.g., because of misfunctioning of traffic control functions caused by unpredictable statistical fluctuations of traffic flows or of network failures. Therefore, additionally, functions referred to as congestion control functions in this Recommendation are intended to react to network congestion in order to minimize its intensity, spread and duration.

Overload treatment applies only to traffic flows with no QoS commitments. It refers to a set of functions in the network that can detect and reduce the amount of overload.

9.1 Traffic control functions

The following functions are identified for traffic control:

i) Network resource management.

ii) Admission control.

iii) Parameter control.

iv) Packet marking.

v) Traffic shaping.

vi) Packet scheduling.

9.1.1 Network resource management

Network resource management refers to a set of policies and rules for allocating the network resources such as the bandwidth and buffer spaces in a switch or router.

9.1.2 Admission control

Admission control refers to the policies of the network to admit commitments to a new IP flow or to refuse the commitments when the demand for the network resources such as the bandwidth and buffer spaces exceeds the available capacity in the network.

9.1.3 Parameter control

Parameter control is a set of policies that monitors and controls that the traffic contract is not exceeded.

9.1.4 Packet marking

When an IP flow is found to be non-conforming to one or more aspects of the traffic contract, the non-conforming packets may be marked by modifying the relevant bits in the ToS/DS field of the IP header.

9.1.5 Traffic shaping

Traffic shaping is an action by the network to modify the traffic characteristics of the flow such that the flow becomes more suitable for the network. One example is shaping to the peak rate of a flow.

9.1.6 Packet scheduling

Packet scheduling is a function of the network with two different objectives. One objective is to bind the queuing delay for a flow that has stringent delay requirements. The other objective is to divide an available resource (e.g., bandwidth) among different flows in a network-specific manner.

9.2 Functions for congestion control and overload treatment

The following functions are identified for congestion control and overload treatment:

– packet discard control;

– explicit congestion notification.

9.2.1 Packet discard control

Packet discard control is used to discard marked traffic in a congested situation. Another application of packet discard control is, in a congested situation, to discard packets for which no stringent QoS commitments apply.

9.2.2 Explicit congestion notification

Explicit congestion notification is an optional method with which the network may inform the receiver of existing network congestion. The receiver can use this information to take appropriate action, e.g., to inform the sender to reduce its transmission rate.

10 Methods and tools for IP traffic engineering

10.1 MPLS traffic engineering

A major goal of traffic engineering is to facilitate efficient and reliable network operations while simultaneously optimizing network resource utilization and traffic performance.

In IP networks, MPLS traffic engineering is one of the most popular traffic engineer methods.

In MPLS traffic engineering (MPLS-TE), there are mainly three steps to establish an LSP. In the first step, each node should have the knowledge of current properties (such as bandwidth, administration statutes) of all the links in the network. In order to do it, the current routing protocols (such as IS-IS and OSPF) are extended to advertise these properties. In the second step, CSPF is used to compute the paths complying with certain constraints in each node. After the path is computed, in the third step, the RSVP-TE signalling will be sent along the path as a label distribution protocol. Each node receiving the signalling will update its relevant properties, such as available bandwidth.

With MPLS-TE, the paths satisfying some constraints can be established. The traffic going along the paths can meet its QoS requirements while optimizing the network resources.

Also, the MPLS-TE can be used to improve network reliability. MPLS-TE fast reroute (FRR) can be used to switch traffic from the original path to the backup path in the scale of 50 ms when the original path is broken.

10.2 DiffServ-aware MPLS traffic engineering

MPLS-TE operates on an aggregate basis across all DiffServ classes of service. DiffServ-aware MPSL-TE (DS-TE) performs traffic engineering at a per-class level, instead of at an aggregate level.

In DS-TE, there are eight class types. Routing protocols, path computation and path signalling have to be enhanced to handle the class types. In the routing protocols, the available bandwidth for each class type is advertised. In path computation, CSPF is enhanced to compute paths with the class type information. Finally, in the RSVP-TE, the class type information is sent to each node in the path. For a DS-TE LSP, it is used only with one class type.

With DS-TE, fine-grained optimization of transmission resources and further enhanced network performance and efficiency can be achieved.

Annex A

The token bucket and the generic byte rate algorithm

(This annex forms an integral part of this Recommendation)

This annex provides several algorithms, which are used to define packet conformance of an IP flow with respect to a rate R (in bytes per second) assuming a bucket size B. The algorithms are considered equivalent in the sense that they will identify the same packets to be conforming.

Traffic metering is a basic component of packet networks that provide various levels of quality of service, such as throughput guarantees and upper bounds on delay and loss. An IP packet stream must conform to a predefined traffic profile in order to assure that the network resources allocated to that stream can provide the expected service level, and to assure that this stream does not cause other network traffic to experience an unacceptable service level.

The starting point is the token bucket. This algorithm is considered equivalent to the continuous state token bucket. This algorithm is shown to be equivalent to the generic byte rate algorithm (GBRA). The GBRA defines the characteristics that relate an information rate parameter to a burst tolerance parameter. Together, these two traffic parameters and their interrelationship form the basis for a generalized traffic profile conformance test. A secondary goal of GBRA is to avoid referencing one or more implementation options on the basis that it is not central to the conformance concept and that it is a source of unnecessary bias. GBRA specifies the fundamental aspects of a conformance test to a single profile, and as such does not include other related policing functions such as multi-stage metering, packet colouring, packet marking, and packet discard policies.

A.1 The token bucket

This algorithm is described as follows (compare also e.g., [b-IETF RFC 2698]):

The token bucket has two fixed parameters:

– the token bucket rate, R, (in bytes per second);

– the token bucket size, B, (in bytes).

The token bucket uses the variable:

– token count, T_c, (in bytes) which is initially (at the arrival at time t_a of the first packet of the flow) full, i.e., the token count $T_c(t_a) = B$.

Thereafter, the token count T_c is incremented by one every $1/R$ seconds, up to B.

When a packet of size N bytes arrives at time t_a, the following happens:

– If $T_c(t_a) - N < 0$, the packet is non-conforming; else

– the packet is conforming and T_c is decremented by N.

NOTE – The arrival time is considered to be when the last bit of the packet arrives.

A.2 The continuous-state token bucket

To be more precise and unambiguous, a continuous-state token bucket is proposed. The continuous-state token bucket is considered to be equivalent to the token bucket. The continuous-state token bucket has the advantage that it is easily implementable.

The continuous-state token bucket has two fixed parameters per IP flow:

– the token bucket rate, R, (in bytes per second) for the flow;

– the token bucket size, B, (in bytes) for the flow.

The continuous-state token bucket uses the following variables:

– the token count T_c (in bytes) of the flow;

– LCT, which is the last conformance time of the flow (in seconds).

Initially (at time t_a of the arrival of the first packet of the flow):

– $T_c = B$;

– LCT = t_a.

At arrival of a packet with size N (bytes) at time t_a:

$T_c' = \min(T_c + R*(t_a - \text{LCT}), B)$

If $T_c' < N$

then the packet is not conforming;

else the packet is conforming

$T_c = \min(T_c', B) - N$

For the continuous-state token bucket, the variables T_c and LCT are only modified at packet arrival.

NOTE – A packet can never be conforming if $N > B$.

A.3 The generic byte rate algorithm

There are two variants for the generic byte rate algorithm: one is the virtual scheduling byte rate algorithm and the other is the continuous-state leaky bucket byte rate algorithm.

For both algorithms, the fixed parameters per IP flow are:

– The increment, I, in units of seconds per byte. I is related to R by $I = 1/R$.

– The limit, L, in seconds. L is related to R and B by $L = B/R$.

A.4 The virtual scheduling byte rate algorithm

The virtual scheduling byte rate algorithm uses the following variable:

– TAT, in units of seconds, is the theoretical arrival time of the next packet.

Initially (at time t_a of the arrival of the first packet of the flow):

– TAT = t_a

The theoretical arrival time (TAT) is updated only after a packet is considered conforming by the algorithm.

At arrival of a packet with size N (bytes) at time t_a:

If $t_a < \text{TAT} + N*I - L$

then the packet is not conforming;

else the packet is conforming

$\text{TAT} = \max(t_a, \text{TAT}) + N*I$

The conformity of the packet is assessed after taking into account the effect of its packet size, N. Indeed, the test uses the previous computed TAT to which N*I is added. In case the packet is not conforming, TAT is not updated.

Therefore, conformity to the virtual scheduling byte rate algorithm may be referred to as a posterior-conformity.

A.5 The continuous-state leaky bucket byte rate algorithm

The continuous-state leaky bucket byte rate algorithm uses the following variables:

– LCT is the last conformance time of the flow (in seconds).

– X is the leaky bucket counter (in seconds).

Initially (at time t_a of the arrival of the first packet of the flow):

– LCT = t_a;

– $X = 0$

At arrival of a packet with size N (bytes) at time t_a:

$X' = X - (t_a - \text{LCT})$

If $X' + N*I > L$

Then the packet is not conforming;

Else the packet is conforming

$X = \max(0, X') + N*I$

As for the virtual scheduling byte rate algorithm, the conformity of the packet is assessed after taking into account the effect of its packet size, N. Indeed, the test uses the computed X' to which $N*I$ is added. In case the packet is not conforming, X is not updated.

Therefore, conformity to the continuous-state leaky bucket byte rate algorithm may be referred to as a posterior-conformity.

NOTE 1 – Equivalence of the algorithms:

– The continuous-state leaky bucket byte rate algorithm is equivalent to the continuous-state token bucket: Just set $T_c = (L–X)*R$, $T_c' = (L–X')*R$ and use $B = L*R$.

– The virtual scheduling byte rate algorithm is equivalent to the continuous-state leaky bucket byte rate algorithm: Just observe that after the two algorithms have been executed, one has TAT = X + LCT.

NOTE 2 – The relationship between the GBRA and the GCRA in [ITU-T I.371]:

In case of ATM, the cell size is N = 53 bytes and $N*I$ becomes equal to T, which is the inverse of the cell rate. If one sets $L = T + \tau$ where τ is the tolerance of the GCRA, then one sees that the GCRA is a special case of the GBRA (see Annex A of [ITU-T I.371]).

It has to be noticed that in the ATM context, in view of the difference of the definitions given above, the conformity of cells is assessed before taking into account the effect of the cell size.

A.6 Throughput behaviour of the GBRA algorithm

In the ATM context, Appendix III of [ITU-T I.371] (Throughput behaviour of the generic cell rate algorithm (GCRA)) describes an unexpected cell discard phenomenon when applying the reference conformance testing algorithm, or GCRA, to a CBR flow that slightly exceeds its agreed upon PCR. This unexpected cell discard phenomenon can only be observed if the CDV tolerance, τ, used in the GCRA is chosen smaller than the peak emission interval, T_c, used in the GCRA. This situation is infrequent since usual values of tolerance τ would lead to low bit rates. Indeed this unexpected situation is avoided if:

$$\tau > T_c \Leftrightarrow PCR(cells/s) = \frac{1}{T_c} > \frac{1}{\tau}$$

As examples, if τ = 200 µs, then PCR > 5000 cell/s (212 kbit/s); if τ = 15 ms, then PCR > 67 cell/s (28 kbit/s).

Since the GBRA is derived straightforwardly from the GCRA used in ATM, the above observation applies. Therefore, when computing a value for I, care should be taken that:

$$L > I$$

In practical situations, the packet delay variation tolerance or limit L should be set to a value relevant for the application under measurement. For example if an MPEG-2 video flow is observed, the limit L should relate to the behaviour of the playout buffer used in the reception side (set-top box in the ADSL case). In the case of a playout buffer implementing an initial delay corresponding to half the buffer, the limit L would be equal to the other half. Classical implementations provide buffer sizes of 80 ms, thus leading to a value of L=40 ms. According to the commonly used UDP/IP encapsulation, MPEG flows are carried in 1344-byte IP frames. Then, the minimum rate value that can be provided by the GBRA algorithm is 268'800 bps.

Annex B

Behaviour of two coordinated generic byte rate algorithms

(This annex forms an integral part of this Recommendation)

The conformance definition of the statistical bandwidth transfer capability is based, besides the test of the maximum allowed packet size, on two instances of the generic byte rate algorithm (see Annex A) where the two instances operate in a coordinated mode. Coordinated mode means that the states of the GBRA are updated if and only if an arriving packet is conforming to all relevant instances of the GBRA. An IP packet passing this test is conforming when it is at least conforming to both the peak GBRA(*Rp*,*Bp*) and the sustainable GBRA(*Rs*,*Bs*).

Figure B.1 shows the behaviour that results from two instances of the GBRA operated in a coordinated mode.

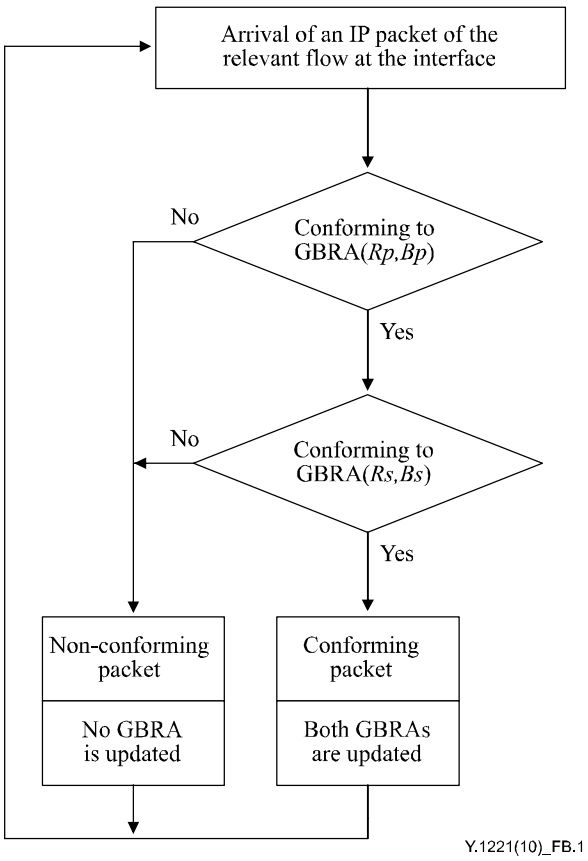


Figure B.1 – Behaviour of two coordinated generic byte rate algorithms

Appendix I

Illustration of token bucket behaviour

(This appendix does not form an integral part of this Recommendation)

Introduction – Token bucket behaviour

A rate, expressed in bytes/s, may be used to partially characterize a flow. The value of the rate may be dependent on the period over which it is measured. For example, a source that sends 50 bytes every 1 ms is different from a source that sends 500 bytes every 10 ms, though their long-term average rates are the same. To allow this difference to be reflected, a token bucket (see Annex A) with bucket rate *R* and a bucket size *B* is used to characterize the flow. In practice, real flows may not show the same recurrent pattern as in the examples above. Yet, the token bucket characterization may be used to characterize the 'worst-case behaviour' of all possible flows of which all packets are conforming to a token bucket with given parameters (*R*,*B*). This will be further described and illustrated in the following.

I.1 Characterization of a flow with a single rate and a single token bucket

A flow is said to be confined or upper bounded by a token bucket with rate *R* and token bucket size *B* if all packets of the flow conform to the token bucket algorithm with bucket rate *R* and bucket size *B* (see Annex A).

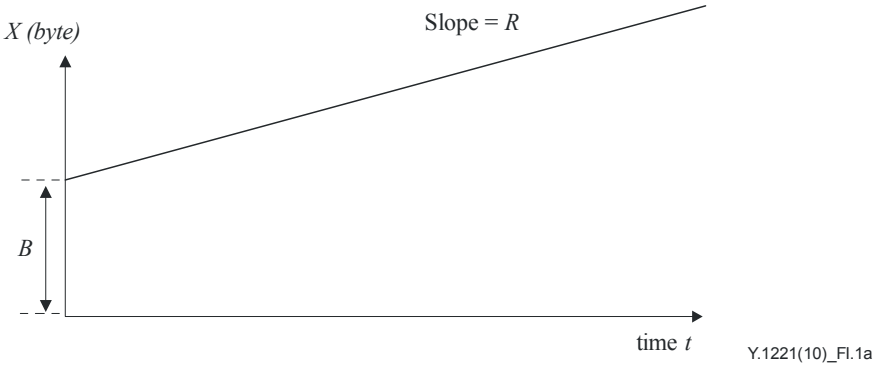


Figure I.1a – Single rate, single bucket size token bucket – 'worst-case' flow

Figure I.1a shows the behaviour of a theoretical worst-case flow which conforms to the token bucket algorithm with token bucket size *B* and rate *R* and where the line rate is assumed to be infinite. The horizontal axis represents the time since the first packet on that flow; the vertical axis shows the maximum number of bytes, *X*, that could have been sent on the flow and still may pass the token bucket up to a time *t*. This theoretical worst-case flow has sent a burst of *B* bytes (at infinite rate) followed by a constant flow of 1-byte packets each 1/*R* second. Note that in a practical IP network, with a minimum packet size larger than 1 byte, a real flow cannot show such a packet rate. It is easily verified that this is the maximum behaviour allowed by the token bucket and that it is indicated by the top line with slope equal to the token bucket rate, *R* (i.e., $X \leq B + t \times R$).

Another worst-case flow may show a recurrent pattern of *B* bytes (at infinite rate) at regular intervals such that the long-term average rate equals the bucket rate *R*. Figure I.1b indicates the trajectory of this flow as a dashed line. In a practical IP network, a real flow can realize or approximate such behaviour. It is easily verified that, in order to send the second burst of *B* bytes, the flow should show reduced activity (e.g., be silent) for a sufficient number of tokens to be replenished.

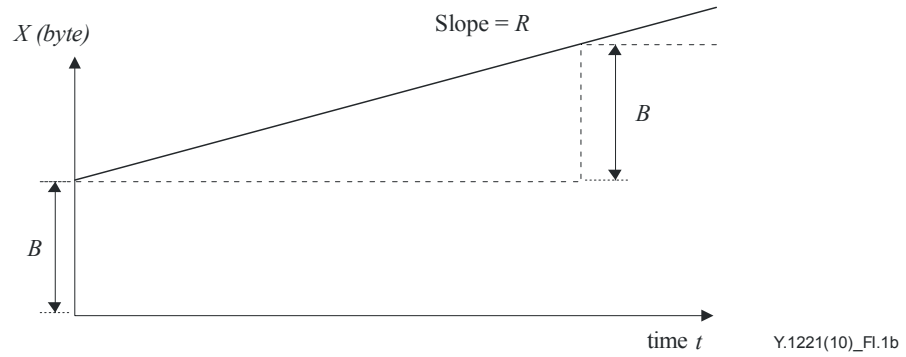


Figure I.1b – Single rate, single bucket size token bucket – 'worst-case' bursty flow

Figure I.1b also illustrates that the expression $X \leq B + t \times R$ still holds but is not at all moments an accurate representation of the actual number of bytes sent on this flow.

In a third example, the flow is less active for some time and then resumes a 'worst-case' behaviour as in the first two examples. This is illustrated in Figure I.1c.

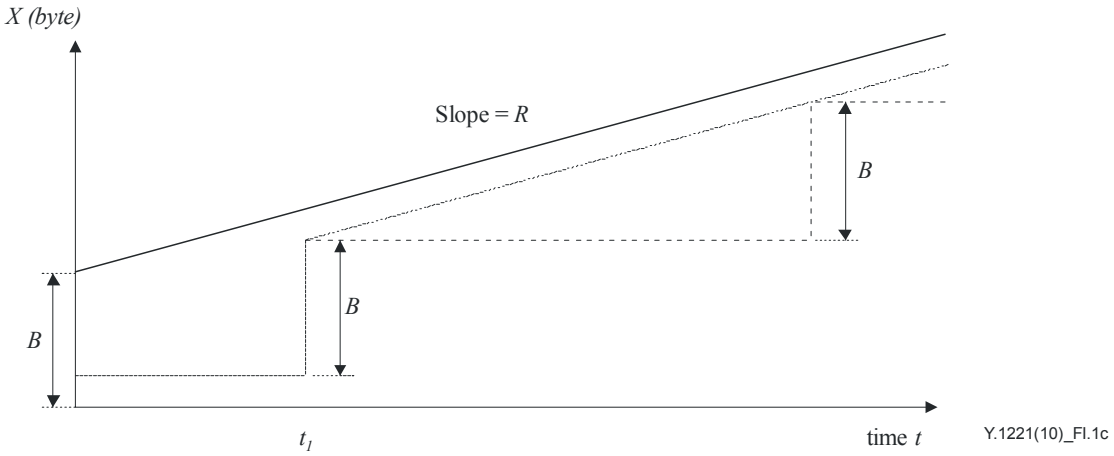


Figure I.1c – Single rate, single bucket size token bucket – 'worst-case' flows after inactivity

The first theoretical worst-case flow resumes maximum activity at time t_1. Due to the modest activity in the pre-t_1 era, the token bucket has been fully replenished to B bytes. This means that, at moment t_1, the flow is constrained to sending a burst not larger than B bytes (at infinite rate) and may subsequently show the (only theoretically possible) worst-case behaviour of sending 1-byte packets each $1/R$ second. Figure I.1c illustrates the trajectory of this flow as a dotted line.

A second worst-case flow, which resumes maximum activity at time t_1, shows a recurrent pattern of B bytes (at infinite rate). This flow also finds itself constrained by the token bucket size B and, when all tokens have been consumed, needs to wait at least B/R seconds before sending the next B-size packets. Figure I.1c illustrates the trajectory of this flow as a dashed line.

Real flows may show many different characteristics, for example by sending different-size packets and by sending at different rates during the lifetime of the flow.

It is the token bucket algorithm that confines the flow, at any moment, in two ways:

– it confines the flow to a burst size (not exceeding B bytes); and

– it confines the flow to a rate (R bytes/s) measured over a sufficiently long period (not exceeding B/R seconds) and disregarding the allowed burst size.

Given a suitable set of token bucket parameters, the network can rely on these simplified flow characteristics for its resource allocation, instead of requiring the originator of the flow to specify its real (possibly very complicated and possibly very diverse) flow or source characteristics in order to take them into account. For a given flow, there may be more than one set of token bucket parameter sets (R,B) to characterize the flow. It is the user's responsibility to select a suitable parameter set for his flow.

NOTE – It should be noted that, after a sufficiently long period of reduced activity, the expression $X \leq B + t \times R$ still holds, but it is no longer an accurate representation of the maximum number of bytes that can be sent on the flow.

I.2 Characterization of a flow with two rates and two token buckets (*Bp* < *Bs*)

When two rates are defined for a flow, then two token buckets are used. The higher of the two rates is designated as the peak rate (*Rp*), the lower rate is the sustainable rate (*Rs*). When the token bucket size for the peak rate is smaller than the token bucket size for the sustainable rate (as is the case for IETF GS and CLS), then the peak rate represents the short-term throughput. The sustainable rate (*Rs*) represents the long-term throughput.

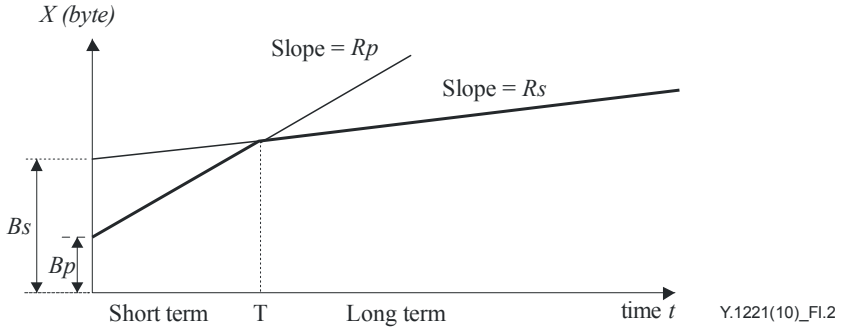


Figure I.2 – Two rates, two token buckets (*Bp* < *Bs*) – 'worst-case' flow

Figure I.2 shows the behaviour of a theoretical worst-case flow which conforms to two token buckets where one has rate *Rp* and token bucket size *Bp*, and the other has rate *Rs* and token bucket size *Bs*, and $Bp < Bs$. The line rate is assumed to be infinite. The number of bytes X is the maximum number of bytes that could have been sent on the flow as a function of time, t, since the first packet arrival, and still pass both token buckets. The number of X is determined by the lines whose slopes are equal to the token bucket rates *Rp* and *Rs*. At times before T the higher rate (i.e., *Rp*) token bucket dominates the worst-case flow behaviour while for time values exceeding T the lower rate (i.e., *Rs*) token bucket dominates the worst-case flow behaviour. These regions are indicated as short term and long term in Figure I.2.

The following relationships define the maximum number for X as a function of time (t), and are indicated with thick lines in Figure I.2.

$$X \leq Bp + Rp \times t \qquad (0 \leq t \leq T)$$

$$X < Bs + Rs \times t \qquad (t \geq T)$$

As in the case described in clause I.1, a suitable set of token bucket parameters may be used to better characterize some flows than would be possible with a single-rate token bucket, and the network can rely on these flow characteristics for its resource allocation.

NOTE – As in the case described in clause I.1, it should be noted that, after a sufficiently long period of reduced activity, the expressions above still hold but they are no longer an accurate representation of the maximum number of bytes that can be sent on the flow.

I.3 Characterization of a flow with two rates and two token buckets (*Bp* > *Bs*)

When two rates are used and the token bucket size for the peak rate is bigger than the token bucket size for the lower rate (as is the case for IETF Diffserv), then the two profile lines that are shown in Figure I.2 do not cross each other anymore. The traffic descriptor includes (*Rp*,*Bp*) and (*Rs*,*Bs*), and defines a committed rate token bucket TB(*Rs*,*Bs*), in which *Rs* is the lower rate, which represents the traffic profile below which the network guarantees delivery of the packets. Packets that conform to this TB are conforming packets. This is very similar to the case of conforming packets for the single rate characterization in clause I.1. For the sake of discussion, these conforming packets are also indicated as "green" packets.

The traffic descriptor also defines a peak token bucket TB(*Rp*,*Bp*), in which *Rp* is the higher rate that represents the traffic profile above which the network does not provide any delivery guarantee. This allows to further distinguish between packets that are not conforming to the sustainable TB.

A packet which is non-conforming to the peak rate TB is called a "red" packet. This is very similar to the case of a non-conforming packet of the single rate characterization in clause I.1, yet with different token bucket parameter set.

A packet, which does not conform to the lower bucket but does conform to the peak rate bucket, is called "yellow". The network may offer different 'commitments' to these non-conforming (i.e., yellow and red) packets. For example, for yellow packets a lesser commitment may apply than for conforming (green) packets and a better commitment than the one that applies to red packets. Figure I.3 illustrates the lines with slopes *Rs* and *Rp*, which separate three different regions as follows.

$X \le Bs + Rs \times t$ → conforming region (green)

$X > Bp + Rp \times t$ → non-conforming region (red)

$Bs + Rs \times t < X \le Bp + Rp \times t$ → non-conforming region (yellow)

NOTE – Clarification on the use of the distinction between the two non-conforming regions (yellow and red) is requested.

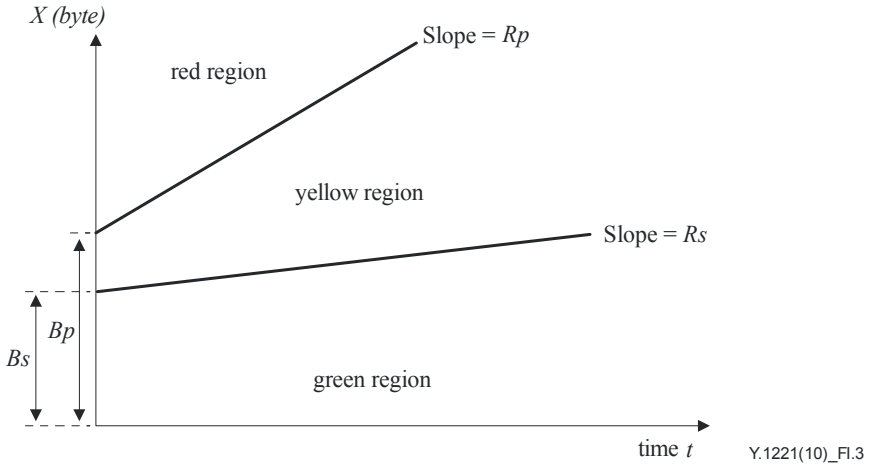


Figure I.3 – Two rates, two token bucket (*Bp* > *Bs*) – 'worst-case' flow

Describing the possible characteristics of a 'worst-case' flow in this configuration becomes fuzzier.

As for the worst-case flows described in clause I.1, a 'worst-case green' flow can be described in which all packets conform to the TB(*Rs*,*Bs*).

Also, a 'worst-case yellow' flow can be described in which all packets conform to the TB(*Rp*,*Bp*). Note that a 'worst-case yellow' flow consists not only of yellow packets but of a mixture of green and yellow packets.

Finally, a flow which is not conforming to TB(*Rp*,*Bp*) will consist of a mixture of red, yellow and green packets, which depends on the severity of exceedence of the respective token buckets.

Appendix II

Relation between IPTC/QoS class and IETF IntServ/DiffServ-specification

(This appendix does not form an integral part of this Recommendation)

The main body of this Recommendation describes two IP transfer capabilities (see clauses 8.1 and 8.2) which strive for compatibility with service descriptions and specifications generated by IETF IntServ and DiffServ groups. The intention is that network elements and networks conforming to these specifications, are likely able or adaptable to support the IP transfer capabilities specified in this Recommendation.

This informative appendix lists and motivates the most important differences between the IP transfer capabilities and the IETF counterparts.

II.1 Dedicated bandwidth IPTC associated with a suitable QoS class

The description for the DBW IP transfer capability (see clause 8.1.1) lists that the DBW capability strives for compatibility with the guaranteed service (GS) [b-IETF RFC 2212] and the end-to-end services based on the expedited forwarding per-hop behaviour (EF-PHB) [b-IETF RFC 2598].

Because ITU-T uses a combination of an IPTC and a QoS class, the following assumes that a QoS class with specified loss commitments and specified delay (variation) commitments is selected for a given flow. A flow characterized by the DBW capability in combination with such a QoS class is in the sequel referred to as 'DBW' . Similarly, flows which use the GS or the EF PHB are referred to as 'GS' and 'EF', respectively.

The common property between DBW, GS and EF is that the network commits to transport IP packets with a specified capacity (bytes/s) with a low delay and low delay variation.

The following differences are noted:

– The DBW capability allows the network provider to discard (possibly all) non-conforming packets, whereas in GS and EF non-conforming packets are expected to be carried on a best-effort basis.

  NOTE – This text description may result in different characteristics specified for DiffServ EF specified in [b-IETF RFC 3246]. In clauses 2.6 and 2.10 of [b-IETF RFC 3246] reorder of a packet in a microflow is not allowed, while putting non-conforming packets into a different class, such as best-effort class, may result in a packet reordering.

  It is expected that the ability to restrict the influx of DBW packets to the agreed rate and characterized by GBRA(*Rp*,*Bp*), greatly simplifies the traffic control in the network (e.g., priority queuing) and minimizes queuing delays for all conforming traffic. It is further expected that most applications using DBW (e.g., streaming media) are able to specify suitable parameters and generate traffic within these bounds (conforming packets).

– The association with a QoS class allows the end-to-end commitments to be *a priori* (before agreeing on a traffic contract) known to the user. GS allows (e.g., via the RSVP protocol) to inform the user about the maximum queuing delay expected to be experienced on the particular flow path. This information is provided *a posteriori* (after the reservation has been made). In addition, it has been shown that on a path with many hops, the accumulation of per-hop maximum queuing delay values leads to much larger values than the 'maximum delay' (or a suitable quantile thereof) experienced on that path. Consequently, the accumulated information is of little or no practical value to the user. EF does not specify an upper bound to the delay (variation).

– DBW applies (in addition to the maximum packet size) the minimum traffic characterization with a single set of token bucket parameters (*Rp*,*Bp*). GS requires the specification of an additional peak rate parameter; however, without the possibility to specify a 1-point packet delay variation tolerance related to that rate. In addition, GS is invoked using two sets of traffic parameters (Tspec and Rspec, which may have different values) plus a so-called 'slack term'.

II.2 Statistical bandwidth IPTC associated with a suitable QoS class

The description for the SBW IP transfer capability (see clause 8.2.1) lists that the SBW capability strives for compatibility with the controlled-load network element service (CL) [b-IETF RFC 2211] and the end-to-end services based on the assured forwarding per-hop behaviour (AF-PHB) [b-IETF RFC 2597].

Because ITU-T uses a combination of an IPTC and a QoS class, the following assumes that a QoS class with specified loss commitments and possibly unspecified delay (variation) commitments is selected for a given flow. A flow characterized by the SBW capability in combination with such a QoS class is in the sequel referred to as the 'SBW'. Similarly, flows which use the CL or the AF PHB/PSC are referred to as 'CL' and 'AF', respectively.

The common property between SBW, CL and AF is that the network commits to transport IP packets with at least a specified capacity (bytes/s) and allows additional (excess) traffic to be transported within the limits of the available network resources.

The following differences are noted:

– The SBW capability and AF are intended to support applications which do not have stringent delay requirements. CL does not commit to quantified delay properties but aims at properties 'under unloaded conditions', i.e., a delay variation which is small compared to the minimum delay (e.g., propagation delay). This property is expected to be difficult to realize in combination with the ability to efficiently support excess traffic beyond and in addition to the guaranteed rate.

– SBW applies (in addition to the maximum packet size) a double set of token bucket parameters (*Rp*,*Bp*) and (*Rs*,*Bs*). CL also requires the specification of the peak rate parameter, however without the possibility to specify a 1-point packet delay variation tolerance related to that rate. In addition, CL is invoked using two sets of traffic parameters, Tspec and Rspec, which may have different values.

– Currently, the SBW capability distinguishes conforming and non-conforming packets. AF distinguishes three so-called 'drop precedence', yet is required to support at minimum two different drop probabilities.

Appendix III

Guidelines for support of services using IP transfer capabilities in a differentiated services environment

(This appendix does not form an integral part of this Recommendation)

One or more of the transfer capabilities defined in this Recommendation may need to be supported at network boundaries. The transfer capabilities strive to comply with end-to-end services based on differentiated services per-hop behaviours, so a possible way to meet these transfer capabilities is to use the differentiated services architecture as indicated in [b-IETF RFC 2475]. It should be noted that there may be alternative ways besides the following differentiated services architecture to provide these transfer capabilities.

If the differentiated services architecture is chosen to implement these transfer capabilities, then the following guidelines apply:

III.1 Guidelines applicable to all ITU-T Y.1221 transfer capabilities

1) Packet scheduling would be supported at network boundaries. Guidelines for the packet scheduling function based on each service definition are provided below.

2) User-to-network boundaries would support IP packet classification based on information in the following IP header fields: source + destination address ([b-IETF RFC 791]), DS field ([b-IETF RFC 2474], [b-IETF RFC 3260]), and L4 source + destination port as specified in [b-IETF RFC 768] and [b-IETF RFC 793].

3) All network boundaries would support IP packet classification based on information in the IP DS field as specified in [b-IETF RFC 2474].

4) This Recommendation provides a conformance definition for each of the transfer capabilities it defines. The test for conformance to the traffic descriptors for each transfer capability would be performed at the ingress of network boundaries using a packet control function. Guidelines for the packet control function based on each service definition are provided below.

5) All network boundaries would be capable of both, packet marking and packet discard, based on compliance to the conformance definition at ingress.

III.2 DBW guidelines

When the DiffServ architecture is used to implement the DBW transfer capability, the EF PHB, as described in [b-IETF RFC 3246] could be used at network boundaries.

This clause defines the traffic control guidelines for network boundaries suitable to meet the service definitions of the DBW transfer capability. It assumes the use of the EF PHB at network boundaries.

1) Conformance to the DBW traffic descriptor for network boundaries could be determined by a parameter control function that tests that the arrival conforms to GBRA(Rp,Bp) and that the packet length is less than M. After the test, some (possibly all) of the non-conforming packets may be dropped. Conforming packets will always be transmitted.

2) Network boundaries would support a packet scheduler capable of meeting the requirements of [b-IETF RFC 3246].

III.3 SBW guidelines

When the DiffServ architecture is used to implement the SBW transfer capability, the AF PHB, as described in [b-IETF RFC 2597] could be used at network boundaries.

This clause defines the traffic control requirements at network boundaries suitable to meet the service definitions of the SBW transfer capability. It assumes the use of the AF PHB at network boundaries.

1) Conformance to the SBW traffic descriptor for all network boundaries could be determined by a parameter control function that tests that the arrival conforms to GBRA(*Rp*,*Bp*) and GBRA(*Rs*,*Bs*), and that the packet length is less than *M*. If packets are found non-conforming, then some (possibly all) of those packets may be remarked or dropped. Conforming packets will always be transmitted.
2) Network boundaries would support a packet scheduler capable of meeting the requirements of [b-IETF RFC 2597].
3) Network boundaries would support packet discard control capable of meeting the active queue management requirements[1] of [b-IETF RFC 2597].

III.4 BE guidelines

When the DiffServ architecture is used to implement the BE transfer capability, the default PHB, as described in [b-IETF RFC 2474], would be used at network boundaries.

This clause defines the traffic control guidelines for network boundaries suitable to meet the service definitions of the BE transfer capability. It assumes the use of the default PHB at network boundaries.

1) Conformance to the BE traffic descriptor at network boundaries could be determined by a parameter control function that tests that the packet length is less than *M*. If packets are found non-conforming, then some (possibly all) of those packets may be remarked or dropped.
2) Network boundaries would support a packet scheduler capable of meeting the requirements of the default PHB described in [b-IETF RFC 2474].

[1] Active queue management requirements for the AF PHB are described in section 4 of [b-IETF RFC 2597].

Appendix IV

Example methods for determining token-bucket parameters

(This appendix does not form an integral part of this Recommendation)

The appropriate function and its parameters for determining the token bucket size parameter may depend on several factors, such as the application, media type, and transmission protocol. The network operators/carriers should select the most appropriate method depending on the information available.

This appendix provides a few examples of methods for determining token bucket parameters when the only available information is the bit rate of the IP flow.

To monitor the conformance of an IP packet flow with the token bucket algorithm, two parameters are required:

- bucket rate (R);
- bucket size (B).

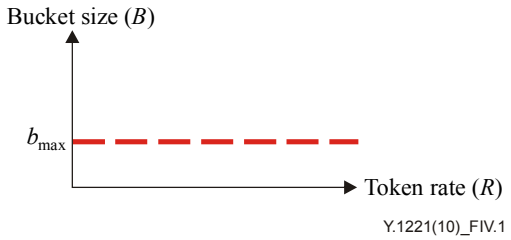
When an application that has an IP-flow requests a QoS guarantee, these two parameters must be specified before a session is established. If these parameters are omitted, the network cannot decide the conformance of each packet, and transmission assurance is not feasible.

However, these parameters are quite often unavailable. For example, when the sessions are established using SIP/SDP, only bucket-rate information is available. The bucket rate is directly related to the bit-rate information of the IP flow, so using SIP/SDP bandwidth information (b=) is a straightforward method. On the other hand, there is no information related to the burstiness of the IP flow, so determining the bucket size parameter is difficult. In general, bit-rate information is easy to know because many application/protocol designers are aware of the bit-rate. However, burstiness is not as easy to know because application/protocol designers often do not care about the burstiness.

Therefore, a method to determine the bucket-size parameter (B) under such conditions is needed.

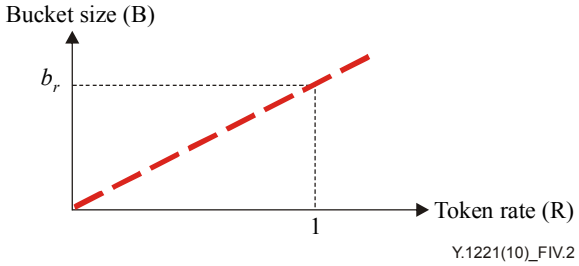
The bucket size, parameter B, corresponds to the burstiness of the IP flow. The value of parameter B may depend on the type of application, end-terminal performance, and interface condition between the user and the network, for example. A few example methods for determining the value of B are given below.

Example 1: $B = b_{max}$ (constant)



When the possible range of the burstiness of all IP flows is known, the maximum bucket size $B = b_{max}$ can be applied to all the flows. In this case, the network does not necessarily need to know the burstiness of each IP flow. However, a huge buffer resource is required for preventing the loss of conforming packets even for very low-bit-rate flows.

Example 2: $B = b_r * x$ (proportional to bucket rate)



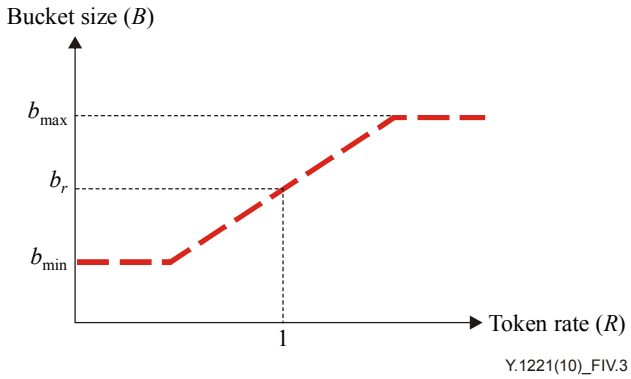
This example is based on the following assumption.

Assumption:

When the bit rate of the IP flow is greater, the burst from the flow becomes greater.

Based on the above assumption, the bucket size of the IP flow is expressed by the non-decreasing function of the bucket rate, b_r. Using a linear function is one of the simplest examples of determining the bucket size.

Example 3: $B = \min(b_{max}, \max(b_r * x, b_{min}))$ (with upper and lower limit)



In the second example, two problems might occur. The bucket size may become too small or too large. For very low-bit-rate flows, the bucket size might become smaller than the size of a single packet. In such a situation, there is a possibility that all packets are discarded as non-conforming packets. On the other hand, for very high-bit-rate flows, the bucket size might become very large; the burst traffic from those flows may cause traffic congestion in the network. To avoid this problem, upper and lower limits can be introduced, mainly from the viewpoint of the implementation.

Bibliography

[b-IETF RFC 768] IETF RFC 768 (1980), *User Datagram Protocol*.

[b-IETF RFC 791] IETF RFC 791 (1981), *Internet Protocol – DARPA Internet Program – Protocol Specification*.

[b-IETF RFC 793] IETF RFC 793 (1981), *Transmission Control Protocol – DARPA Internet Program – Protocol Specification*.

[b-IETF RFC 1633] IETF RFC 1633 (1994), *Integrated Services in the Internet Architecture: an Overview*.

[b-IETF RFC 2210] IETF RFC 2210 (1997), *The Use of RSVP with IETF Integrated Services*.

[b-IETF RFC 2211] IETF RFC 2211 (1997), *Specification of the Controlled-Load Network Element Service.*

[b-IETF RFC 2212] IETF RFC 2212 (1997), *Specification of Guaranteed Quality of Service*.

[b-IETF RFC 2474] IETF RFC 2474 (1998), *Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers*.

[b-IETF RFC 2475] IETF RFC 2475 (1998), *An Architecture for Differentiated Services*.

[b-IETF RFC 2481] IETF RFC 2481 (1999), *A Proposal to add Explicit Congestion Notification (ECN) to IP*.

[b-IETF RFC 2597] IETF RFC 2597 (1999), *Assured Forwarding PHB Group*.

[b-IETF RFC 2598] IETF RFC 2598 (1999), *An Expedited Forwarding PHB*.

[b-IETF RFC 2698] IETF RFC 2698 (1999), *A Two Rate Three Color Marker.*

[b-IETF RFC 3246] IETF RFC 3246 (2002), *An Expeditd Forwarding PHB (Per-Hop Behavior)*.

[b-IETF RFC 3260] IETF RFC 3260 (2002), *New Terminology and Clarifications for Diffserv.*

SERIES OF ITU-T RECOMMENDATIONS

Series A	Organization of the work of ITU-T
Series D	General tariff principles
Series E	Overall network operation, telephone service, service operation and human factors
Series F	Non-telephone telecommunication services
Series G	Transmission systems and media, digital systems and networks
Series H	Audiovisual and multimedia systems
Series I	Integrated services digital network
Series J	Cable networks and transmission of television, sound programme and other multimedia signals
Series K	Protection against interference
Series L	Construction, installation and protection of cables and other elements of outside plant
Series M	Telecommunication management, including TMN and network maintenance
Series N	Maintenance: international sound programme and television transmission circuits
Series O	Specifications of measuring equipment
Series P	Terminals and subjective and objective assessment methods
Series Q	Switching and signalling
Series R	Telegraph transmission
Series S	Telegraph services terminal equipment
Series T	Terminals for telematic services
Series U	Telegraph switching
Series V	Data communication over the telephone network
Series X	Data networks, open system communications and security
Series Y	**Global information infrastructure, Internet protocol aspects and next-generation networks**
Series Z	Languages and general software aspects for telecommunication systems

Printed in Switzerland
Geneva, 2010