Document Number: bs_tmgt_01

IEEE 802.17 Traffic Management Draft

Bob Sultan (bob.sultan@fnc.fujitsu.com)

September 2, 2001

Abstract: This document is a draft of a section describing *traffic management*.

1. Traffic Management

Traffic management defines the rules applied to traffic on a ringlet in support of service guarantees (QoS) and relative service priority (CoS).

1.1 Overview

This draft describes a TM scheme having guaranteed-rate and best-effort traffic classes. The guaranteed-rate traffic class provides a low delay-bound, suitable for voice and other applications with a requirement for low delay. It supports one service class for traffic that is shaped prior to ingress (as in the case of a circuit emulation application) and one for traffic that has well-defined burst characteristics. The best-effort traffic class allows specification of a minimum rate but not a delay bound.

> Author's Note: This document suggests that it is not necessary to *choose* between schemes requiring significant transit queuing (e.g. Cisco SRP) and schemes requiring limited transit queuing (e.g. Nortel iPT) for best-effort (or actually minimum-rate + excess-rate) traffic. These styles can coexist on the ring without significant technical compromise. Both deploy ingress rate control, detect and react to congestion, and are able to limit the scope of congestion control to that portion of the ring where it is needed. The key difference between the styles is that one detects congestion by examining the length of a transit queue and the other detects congestion by detecting the time that a frame waits to gain access to the ring. This document takes the view that a *single* scheme can provide fair access for best-effort traffic, with somewhat different rules for different transit queue depths.
>
> This scheme is likely to have holes, errors, etc., so please let me know what they are. I would be interested in joining with others to form an ad hoc to document a method of Traffic Management that would be generally acceptable.
>
> This document is a draft, has not been reviewed, and is known to contains errors (all the usual disclaimers).

1.2 MAC Reference Model

A station is said to have an ingress, transit, or egress role with respect to a frame, depending on whether the frame enters the station from a MAC-client, passes-through the station on a path from source to destination, or exits the station towards a MAC-client. In the case of the transit and egress roles, TM activities are identified as transit and egress, respectively. In the case of the ingress station role, distinct ingress and access sub-layers are identified, illustrated in Figure 1.1. Thus, there are four rule domains to which rules can be applied; ingress, access, transit, and egress. Rules associated with the ingress and egress demains are identified by *service-class*. Rules associated with the access and transit domains are identified by *traffic-class*.

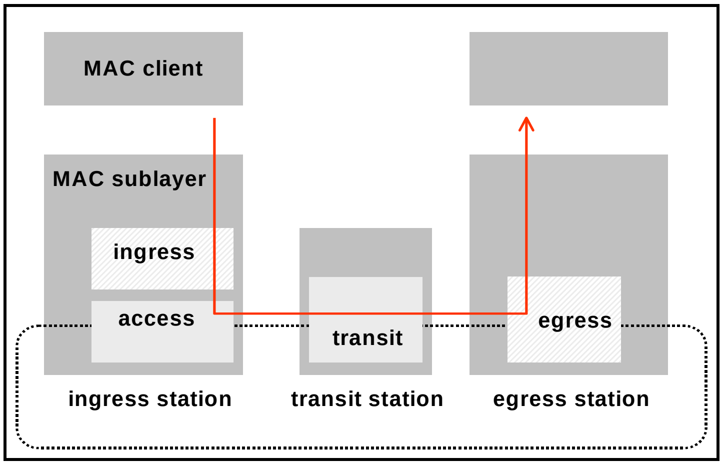

Figure 1.1 – Station Roles and Traffic Management Domains: Ingress and egress domains identify rules by *service-class*. Access and transit domains identify rules by *traffic-class*.

1.3 Service Classes

Each MAC frame is associated with a *service-class* specified at the MAC service interface. The service class identifies rules applied to frames of that class. The service-classes are described in Table B.1.

category	class	rate-commitment	worst-case delay[1]	description
guaranteed-rate	GR1	CIR > 0; Bc = 0	2N x MTU-time[2] (N = maximum stations on ring)	Guaranteed-rate with traffic shaped on ingress; least end-to-end delay.
	GR2	CIR > 0; Bc ≥ 0	2N x MTU-time + Bc/CIR[3]	Guaranteed-rate with no ingress shaping requirement; bounded end-to-end delay.
best-effort	BE	CIR ≥ 0; Bc unspecified	unspecified	Best-effort with optional minimum rate; excess above minimum rate determined by fairness algorithm.

Table B.1– Service Classes

Two guaranteed-rate (GR) service-classes and one best-effort[4] (BE) service-class are described. Both GR service-classes specify a non-zero CIR. GR1 serves traffic that is shaped prior to ingress, as in the case of a MAC-client performing a circuit-emulation application. This service-class provides the least end-to-end delay achievable without segmentation of frames. GR2 permits bursts of size Bc and adds delay necessary to shape traffic towards the ring. Neither GR class supports EIR above the CIR. The BE class supports an ingress-rate, known as the allowed-usage, that can vary over time depending upon ring utilization. The CIR is the provisioned minimum value that can be assumed by allowed-usage. EIR is supported above the CIR and is computed as the difference between the CIR and the allowed-usage.

[1] Approximate ring end-to-end
[2] N x MTU-time for access to ring at ingress plus 1 MTU-time at each transit station
[3] add shaping delay at ingress station to GR1 delay (equal to MBS/CIR)
[4] If a minimum rate is specified, *best-effort* is probably not a good name for this class. Perhaps the categories should be *low-delay* and *unspecified-delay*. …or guaranteed-rate and minimum-rate.

1.4 Traffic Classes

Frames following a specific set of rules for access and transit are identified by a *traffic-class*. One or more service classes can be associated with a traffic-class. Traffic-classes are not visible at the MAC service interface. 802.17 supports guaranteed-rate (GR) and best-effort (BE) traffic-classes. The GR1 and GR2 service-classes are associated with the GR traffic-class. The BE service-class is associated with the BE traffic-class.

1.5 Ingress Rules

Ingress rules reference the model[5] depicted in Figure 1.3. The rules are described 1.5. Ingress rules specify policing and shaping activities.

GR1 traffic is policed to the provisioned GR1-CIR with bursts above the CIR not permitted. That is, the GR1 policer verifies that the GR1 stream was shaped prior to ingress (see section xxx, for shaping conformance rules). GR2 traffic is policed to the provisioned GR2-CIR and GR2-Bc. BE traffic is policed to an allowed-usage value assigned by the BE fairness-algorithm. The BE allowed-usage rate may have a provisioned minimum value. Non-conforming frames of all classes are discarded[6].

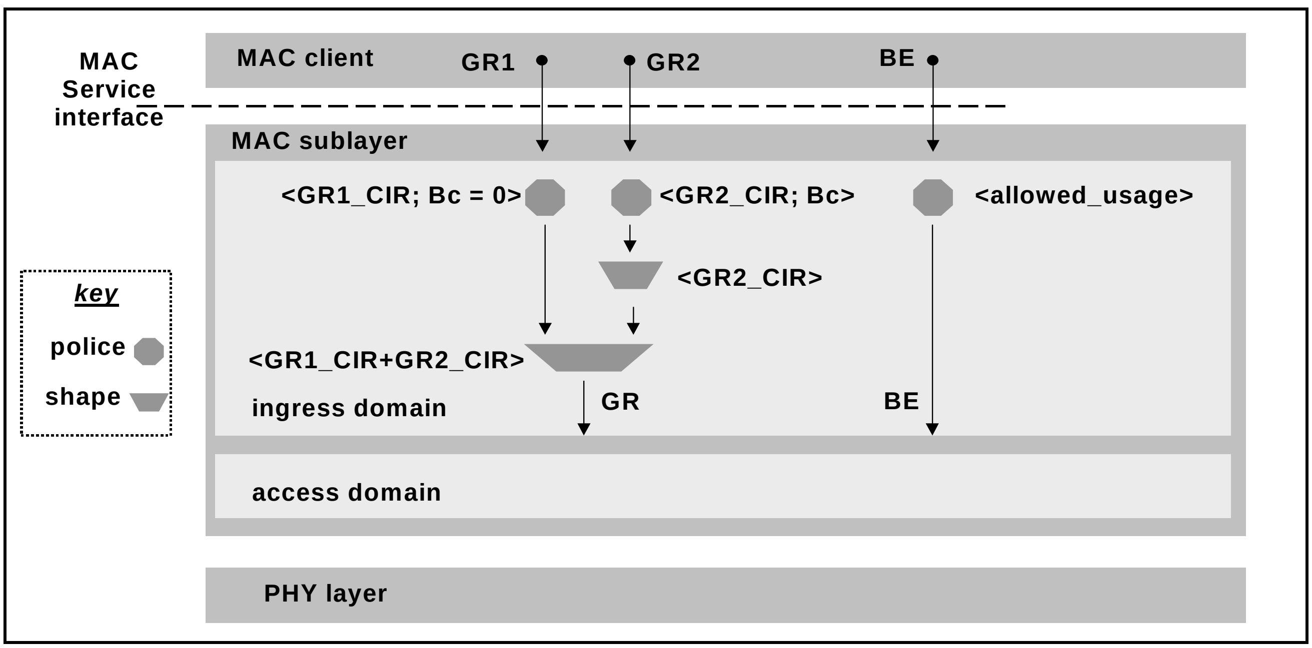

Figure 1.2 – ingress reference model

Frames exiting the GR2 policer enter the GR2 shaper. GR2 traffic is shaped to the GR2-CIR by the GR2 shaper (see section xxx, for shaping conformance rules). Frames exiting the GR1 policer and the GR2 shaper are merged into a single shaped stream of the GR traffic-class by the GR shaper. GR traffic is shaped to the value GR1-CIR + GR2-CIR. Since the GR1 and GR2 streams were individually shaped prior to entering the GR shaper, shaping of the combined stream implies a relatively small buffer (~ 1 MTU) and delay (~1 MTU-time requirement). Frames exiting the GR shaper and the BE policer enter the access domain.

[5] This specific design model need not be implemented in products, but other designs must produce a functionally equivalent result.
[6] Frames are not marked as *discard-eligible* in order to avoid the use of ring capacity by traffic that may later be discarded.

class	policing	shaping	adjustment request from access-domain
GR1	CIR > 0; Bc = 0		
GR2	CIR > 0; Bc ≥ 0	CIR	
BE	allowed-usage ≥ 0; Bc unspecified	unspecified	adjust allowed-usage

Table B.2 – Ingress rules

1.6 Access Rules

Traffic entering the access domain is placed on the GR access-queue or the BE access-queue depending on associated traffic-class. Ingress traffic has priority over transit traffic except for transit frames currently in transmission.

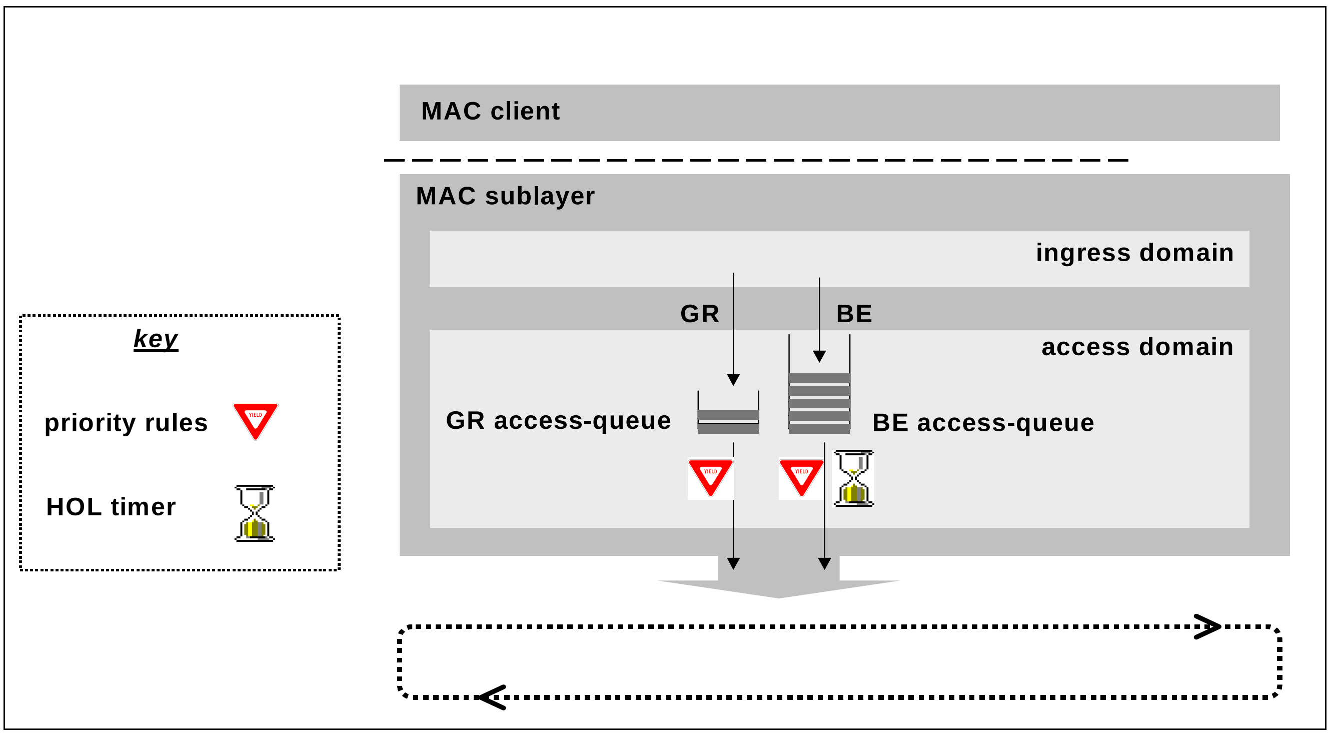

Figure 1.3 – Access reference model

A frame is sent from the head of the BE access-queue when the GR access-queue is empty, no frame is currently *in transmission*, and either there is no BE transit traffic to forward or the BE arbitration algorithm indicates that *ingress* BE traffic should take precedence over *transit* BE traffic. See section xxx, for a description of the arbitration algorithm.

traffic class	access queue depth	transit style	priority	send congestion notification	act on receipt of congestion notification
GR	(N-1) x MTU[7]		yield to transit frame *in transmission*		
BE	unspecified	queued	yield to GR frame (access or transit) yield to transit frame *in transmission* yield to BE frame ready for transit[8] when indicated by arbitration algorithm		transfer request to adjust allowed-usage to ingress-domain
		immediate	yield to complete frame in transmission or ready for transmission	HOL timer expiry	

Table B.3 – Access Rules

1.7 Transit Rules

Transit rules reference the model depicted in Figure 1.4. The rules are described in Table B.4.

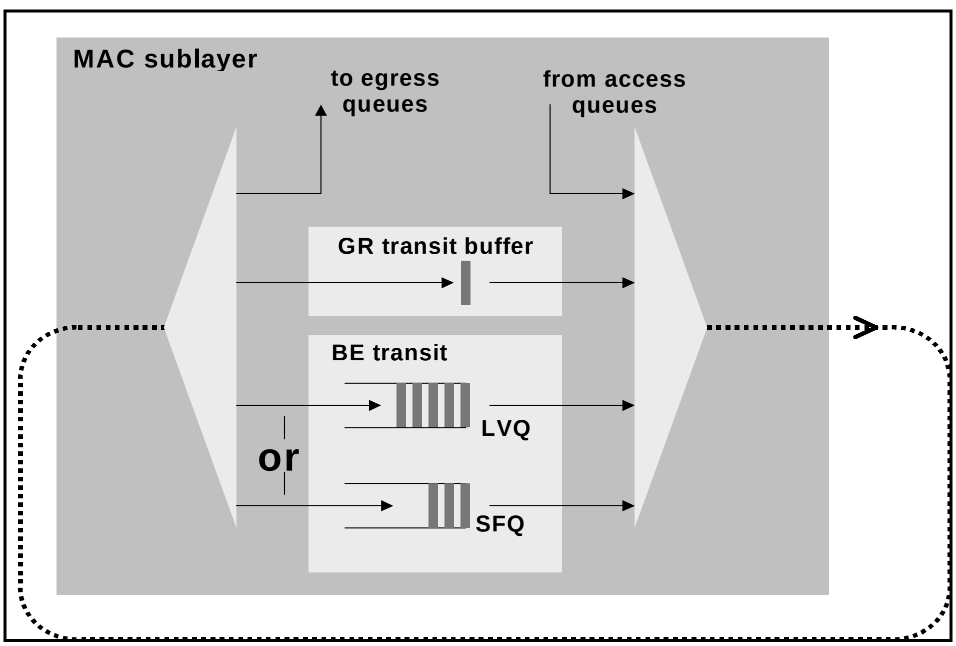

Figure 1.4

[7] To accommodate longest GR transit train
[8] i.e. queued

Transit GR traffic has priority over all other traffic. A GR transit frame is retransmitted as soon as it is fully received and any frame currently in transmission completes transmission.

BE transit can be supported using a *long queue (LQ)* or a *short queue (SQ)*. The *queue-style* is provisioned per-station or inferred from the size of the available BE transit buffer. The two styles of station coexist on the ring, without stations of either style receiving preferential access to resources[9]. SQ BE transit specifies a maximum per-station transit delay of (N+1) x MTU-time. N x MTU-time is allowed for the maximum GR train to pass. An additional MTU-time is allowed for the BE frame itself to transit. This type of transit at a station requires regulation of ingress traffic at the station, since ingress BE traffic must yield to transit traffic. Regulation is provided by adjusting the length of time that a frame is allowed to remain at the head of the BE access-queue waiting for an insertion opportunity. The waiting time is used as a measure of congestion. A shorter time-out period implies a lower threshold for declaring congestion[10].

In the case of LQ BE transit, ingress traffic is allowed to enter the ring even as transit traffic arrives at the station. The transit traffic is absorbed by the LQ until a congestion threshold is reached. Congestion is not determined by the waiting time of the HOL ingress frame, because such frames can take precedence over transit frames, so waiting time does not reflect congestion. Instead, congestion is measured by transit buffer occupancy.

Detection of congestion in a station results in the generation of an upstream congestion notification message as described in RFC 2892. In the case of an SQ style station, congestion is detected via the HOL timer. In the case of the LQ style station, congestion is detected via LQ congestion threshold crossing.

traffic class	**transit queue**	**delay**	**priority**	**congestion trigger**	**action on receipt of congestion notification**
GR		1 MTU-time[11]	yield *only* to frame *in transmission* retransmit after complete frame received		
BE	long variable-length queue (LVQ)		yield to frame in transmission yield to frame(s) in GR access queue yield to frame(s) in BE access-queue when indicated by arbitration algorithm.	threshold on transit queue occupancy	adjust allowed-usage and forward notification as described in RFC 2892 pause retransmission from transit queue for drain-interval[12].
	short fixed-length queue (SFQ)	NxMTU-time	yield *only* to frame *in transmission* retransmit after complete frame received		adjust allowed-usage and forward notification if transit-rate threshold exceeded

[9] In order to insure BE fairness, the two types of congestion thresholds must be set so that neither queued nor immediate-style stations have an advantage in ring access. The relationship could be established by simulation or by other means. This is for further study.

[10] An alternative to the HOL timer for SQ transit is rate measurement. Crossing a threshold on transit rate provides an indication of congestion even when no ingress BE traffic is on the access-queue. This allows corrective action to be taken before frames arrive at the ingress.

[11] Approximate. Implies that frame is transmitted as soon as (1) complete GR transit frame is received and (2)there is no frame in transmission.

[12] Prevents transit queue from further congesting an immediate-style downstream transit station.

Table B.4- Transit Rules

1.8 Egress Rules

Traffic arriving at the ring egress is placed on a queue associated with the service class (GR1, GR2, or BE) for delivery to the MAC client via the MAC service interface.

1.9 Arbitration Algorithm

Arbitration between ingress and transit BE frames for ring access in LQ style stations is described by RFC 1582 (with modifications TBD).

1.10 Allowed-usage Computation

Computation of allowed-usage is as described in RFC 1582 with extensions to compute allowed-usage based on transit rate for SQ stations (details under construction).

1.11 Admission Control

This document will include a 2-phase commitment protocol for distributing information about provisioned GR and BE capacity reservations.

1.12 Overcommitment

This document will describe methods for supporting statistical guarantees for committed-rate service at rates higher than nominal reservations.