2008 Argentine Congress on Computer Science
(Congreso Argentino en Ciencias de la Computación - CACIC 2008)

MPLS Technology: Class of Service

Antonio Ricardo Castro Lechtaler and Rubén J. Fusario
Universidad de Buenos Aires and Departamento Sistemas, Facultad Regional Buenos Aires,
acastro@utn.edu.ar, rfusario@speedy.com.ar

and

Carlos García Garino and Jorge García
Carrera Redes y Telecomunicaciones, ITU and Facultad de Ingeniería
Universidad Nacional de Cuyo; Mendoza, Provincia de Mendoza
cgarcia@itu.uncu.edu.ar, jgarcia@itu.uncu.eduar

and

Alejandro Arroyo Arzubi
Escuela Superior Técnica, Universidad del Ejército
alejandro.arroyo@iese.edu.ar

Abstract

Quality of Service is a sought after feature to handle isochronal signals in data networks. In the cases in which currently used techniques such as best effort are not sufficient, it becomes relevant to analyze Class of Service, since useful conclusions can be drawn from its study. To evaluate the cases under study, we carry out a survey drawing conclusions on the differentiated treatment of flows used by the various techniques under consideration. Among others, we can name: First in First out - FIFO, Priority Queuing - PQ, CoS based queuing, Weighted Fair Queuing - WFQ, Low Latency Queuing - LLQ, Class Based Queuing - CBQ. In addition, we explore Differentiated Services architecture – DiffServ. Finally, we identify the advantages these techniques offer in particular situations.

Keywords: Multi Protocol Label Switching, QoS, Class of Service, Traffic Engineering.

1. INTRODUCTION.

A common way to define Quality of Service in a network is to focus on the actions taken on the data flows to offer a better delivery status.

The evolution undergone in user service requirements forces us to review these concepts to create new definitions. These new definitions must give opportunities for new applications and quality improvements, taking the most advantage of network resources

IP is currently the most widely used network protocol. Considering it is ***not connection oriented***, it has no quality of service. It only offers a service called ***best effort*** which, in practice, does not apply any action on the network flow, nor does it differentiate the different flow types.

In a previous article [1], we discussed MPLS technology [2-6] as a means to provide elements for ***quality of service*** in traditional networks. In this article, we focus on the concept of ***quality of service***, discussed in section 2, and how it can be complemented with other techniques, such as Traffic Engineering [7, 8] and Differentiated Services [9], explored in section 3. One of the focal points of this work, the discussion on the treatment of data flows is presented in section 4. The next section deals with the architecture and implementation of differentiated services. Finally, we draw conclusions from the analysis.

2. QUALITY OF SERVICE - QoS.

The best effort notion is the starting point to provide ***Quality of Service – QoS*** to a data transmission network. Any difference which might yield an improvement will translate in an increase in the quality of the service. From a **classical** viewpoint, to provide QoS concerns the following service features:

- ✓ Guaranteed delivery
- ✓ Transmission error recovery
- ✓ Congestion control
- ✓ Guaranteed bandwidth
- ✓ Flow control

The appearance of new services such as ***Mail***, ***Chat***, ***VoIP***, ***Multimedia,*** and others, added to the traditional data delivery, required the integration of parameters not considered before. For example:

- ✓ **Delay:** Period of time in transit through a path, where the amount of jumps involved becomes a significant parameter.
- ✓ **Jitter:** Difference in the delay through a path, originated by network instability.
- ✓ **Priority:** To ensure even more the adequate delivery of critical traffic.
- ✓ **Availability:** Percentage of time in which service is operative.

At first, the concept of quality of service did not exist. The closest notion involved giving the package its destination address – ***best effort***. Later, the TCP/IP protocol suite made this paradigm more efficient by including detection of lost, mistaken or scrambled packages, and others. It introduced TCP protocol ***time-out*** mechanisms and ***sliding-windows*** to control variations introduced by the network.

With this change of framework, the new objective consists of accomplishing a network where a differentiated treatment can be established for the information in transit, adjusting to software requirements or to a particular service profile associated to a rate.

This was the main reason behind the change in paradigm from ***destination based routing*** to ***Quality of Service Routing – QoSR***. In the latter, ***optimal path*** recognition is based on, in addition to the destination address, some other previous knowledge on network availability, such as ***delay, jitter, package loss levels, and others.***

With this purpose, attention should be paid to a commonly overseen feature, flow differentiation. It becomes necessary to materialize specific methods to identify and mark a particular data flow to grant uniform treatment throughout the path from endpoint to endpoint.

3. OTHER COMPLEMENTARY TECHNIQUES TO PROVIDE QoS.

Currently, other supplemental techniques to the IP protocol exist to provide greater flexibility when requiring quality of service. Those techniques are:

- ✓ **Traffic Engineering - IT** based in the developments accomplished by the MPLS protocol.
- ✓ **Differentiated Services - DiffServ** takes advantage from the information provided in the ToS field of the IP datagram to differentiate flows. In this case, priority makes the difference.
- ✓ **Integrated Services - IntServ** is a third option based in previous commitment of network resources through the entire path, from origin to destination.

Frame Relay as well as ATM provides guaranteed bandwidth, minimum latency, minimum congestion risk, but only the latter provides a differentiated service depending on the type of traffic transported. None of them can guarantee a global solution from one end to the other end of the path.

For Frame Relay, we can define access bandwidth. In doing so, we guarantee ***the Committed Information Rate – CIR***, which results in the bandwidth for each ***Data Link Connection Identifier – DLCI.*** However, the allocation is fixed. If the source consumes less, the exceeding bandwidth quantity cannot be reassigned to another circuit or application.

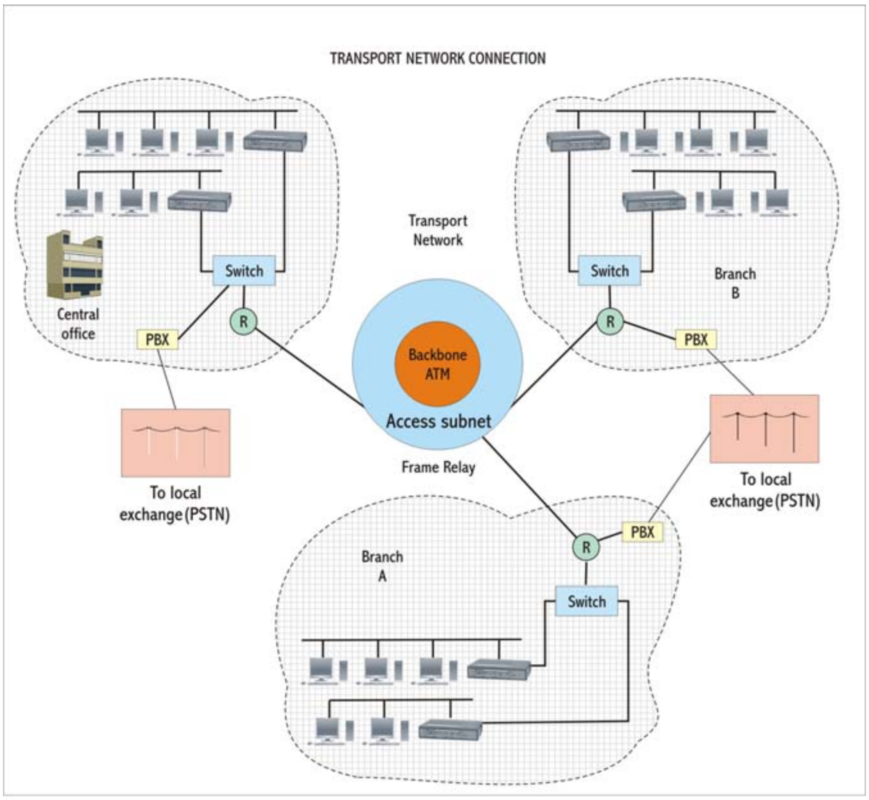

Figure N° 1. Access subnet architecture and transport network backbone

Figure N° 1 shows how ***Frame Relay*** and ***ATM*** often combine to form an access subnet, and a core subnet or backbone respectively, originating networks with greater heterogeneity. These types of situations allow QoS to be controlled only in patches, not granting the possibility for endpoint to endpoint control.

This network architecture is currently implemented in corporate networks. A TELCO provides the transport network to which the different areas of the company log in. However, the main point lies in the bandwidth estimate that each area needs to operate for its needs of software, VoIP, videoconference, central internet access, and others.

Two methods are used to guarantee the absence of bottlenecks at the network level:

- ✓ **Under-subscription**: Expensive to any company.
- ✓ **Over-subscription:** less expensive but may generate congestion.

Thus, more rational and effective methods are required to preventively control the data to avoid network collapse by traffic congestion from a global viewpoint. Among the most widely used techniques we can mention ***token bucket*** and ***leaky bucket.***

In the ***token bucket*** technique, as shown in Figure N° 2, the main objective of the algorithm is to allow the transmission at greater speeds when the source receives a traffic peak. The algorithm works in the following way: the bucket contains tokens generated at a rate **r**. It can allow a maximum of **b** tokens, being at full state when initiated. Recall that each token has one byte.

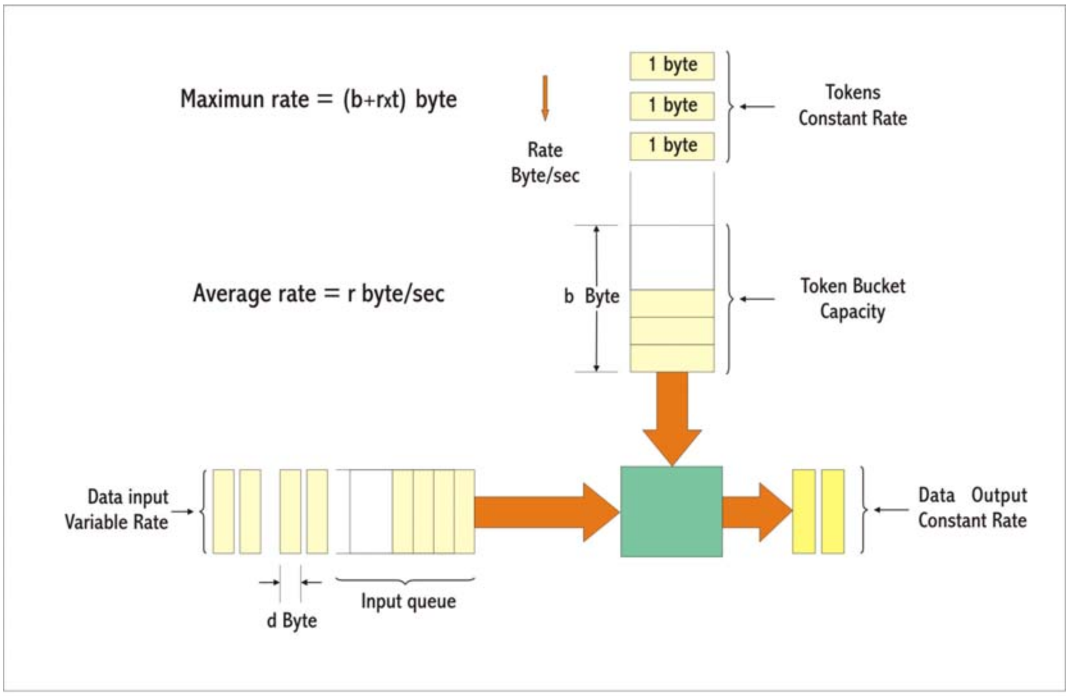

Figure 2: Tocken Bucket model applied to a traffic control network.

To transmit each data byte, a token must be grabbed from the cube and dropped. While tokens are present in the cube, the source can insert traffic in the network at the desired rate. When tokens run out, the network must wait for a new token to be generated. Thus, data transmission decreases at rate **r**.

The parameter **r** establishes the continuous sustainable data rate; while **b** determines the amount by which this rate can be exceeded for short periods of time. In particular, traffic must obey the rule that for any time period, the quantity of sent data cannot exceed **r t + b,** for any interval of **t** time.

At a certain moment, the presence of only one input flow can use up all of the tokens. Thus, this flow will travel at a higher speed than the one assigned individually, but always with a fixed maximum. This method is better suited to ingoing variable flows which must be transformed into maximum controlled flows. Token ingoing rhythm is established by the administration.

Token bucket has a limited capacity. If its capacity is used up, new incoming tokens are discarded and will not be available for future packets. If there are not enough tokens in the bucket to send a packet, the packet must wait until the required tokens are available; otherwise, the packet is discarded.

A ***token bucket*** cannot discard or control priority by itself. Incoming packets to the system cannot be sent immediately as they are queued and delayed in the data buffer.

The ***leaky bucket*** algorithm was introduced by Turner [10]. Since then, it has been widely used to depict network traffic. ***Leaky bucket*** regulates traffic in a "***dripping bucket"*** model as shown in Figure N° 3.

Two parameters are used to describe the algorithm: bucket capacity **s** [bit] and the drainage rate **r** [bps]. The value **r** represents the average transmission rate of the source.

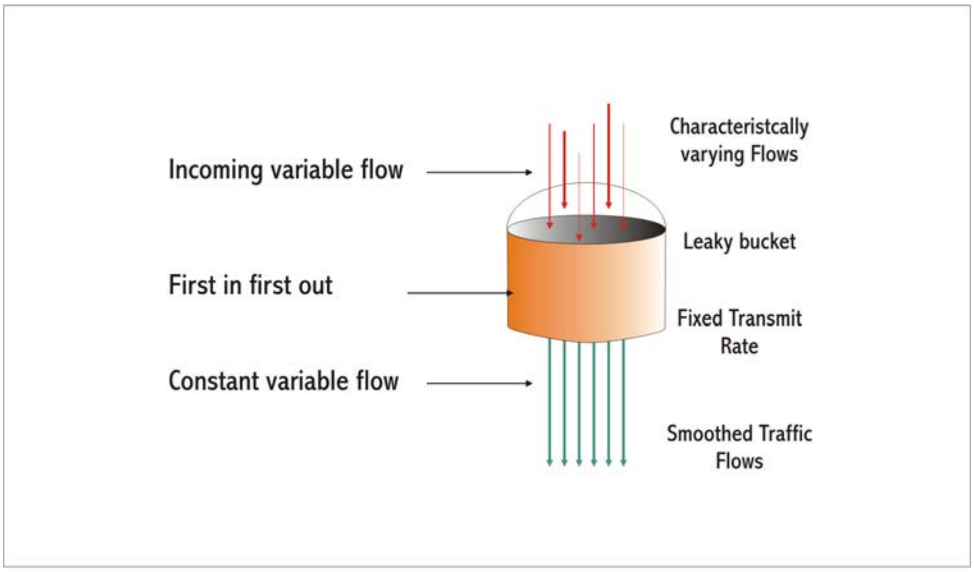

Figure N° 3. Leaky bucket model applied to a traffic control network.

The algorithm works as follows: as long as the bucket contains data, it is sent to the network at rate **r**. All incoming packets go into the bucket. Whenever the bucket is full, the incoming packet loses. In this way, the traffic transmission rate is limited to the **r** value.

The **s** value for a particular flow will be calculated in such a manner as not to lose packets.

The sender should not transmit in the period [0, t] more than **s + r t** bit to avoid packet losses

4. FLOW TREATMENT.

4.1. General Considerations.

Several mechanisms can be used to differentiate flow treatments. They are:

- ✓ Traffic marking techniques
- ✓ Flow differentiation techniques
- ✓ Queue Management
- ✓ Traffic Shaping.
- ✓ Admission Policy.
- ✓ Severe congestion management techniques – RED

4.2. Traffic marking.

Traffic marking is crucial to identify the type of flow of the user application. We will analyze the packet head to establish the type of service required for each IP datagram, and, specifically, the IP packet field called ***precedence***.

The correct bits should be interpreted by traffic control mechanisms at a link level, which should be capable of sustaining the required QoS from endpoint to endpoint. Figure N° 4 details the *precedence* field used inside the IPv4 datagram.

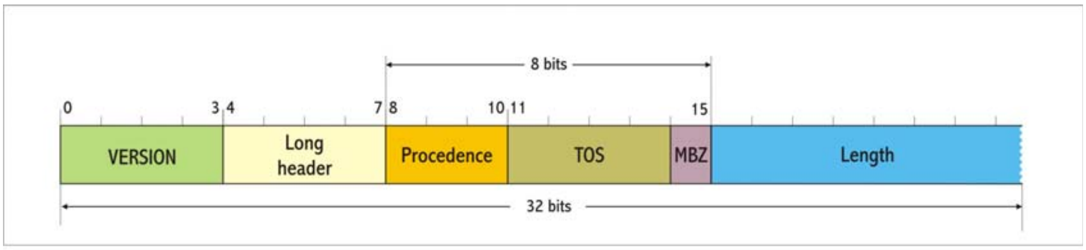

Figure N° 4. Type of Service octet in the IP datagram header Ipv4

The detail of the precedence field can be analyzed with greater precision at the RFC1812 and in the TOS field of the RFC1349

4.3. Flow Differentiation.

From the viewpoint of ***flow differentiation,*** the idea of ***differentiated Class of Service*** or ***differentiation by flow*** involves analyzing and classifying the various types of information flows under the following:

- ✓ Transport and network protocol: IP, UDP, TCP, IPX, DECNET, among others.
- ✓ Origin Port: service requirement
- ✓ Destination Port: service presentation
- ✓ Origin Host: Traffic originator
- ✓ Destination Host: Traffic destination
- ✓ Network Access Interphase: the network shall have an access/departure interphase through which the endpoints will connect.
- ✓ Flow classification at origin: The host at the origin can mark a required type of service.
- ✓ Other definitions based on Access Control Lists (ACLs)

The access interphase shall have, in addition, access policies (admission control) consistent with best network performance criteria. Consequently, only filtered and classified traffic shall enter the network.

An adequate access policy shall have enhanced resource management, providing the ability to stabilize its parameters within reliable performance margins. This condition is essential for QoS stability over time.

4.4. Queue Management.

Queue Management makes reference to the process by which a network device receives traffic from an access interphase and places it in an outgoing interphase, waiting for delivery. This technique is known as ***queue management process*** or ***store-and-forward***. It also controls the queue length in each interphase, handling latency times and congestion notices with various degrees of warning.

Among the most commonly used algorithms we can mention:

- ✓ First in First out – FIFO.
- ✓ Priority Queuing - PQ.
- ✓ CoS Based queuing
- ✓ Weighted Fair Queuing - WFQ.
- ✓ Low Latency Queuing - LLQ.
- ✓ Class Based Queuing - CBQ.

Their main features are:

- ➢ **First in First out - FIFO*:*** it is an efficient standard method as long as the queue does not grow too long. When the queue is complete, packets are dropped.

- ➢ **Priority Queuing - PQ*:*** it involves queue rearrangement following the priority status indicated in the traffic type. It is a slow process involving some packet losses.

- ➢ **CoS Based Queuing:** it is a variation of priority queuing. Several queues are defined and the packets are arranged following the requested type of service. Queue management will depend on each service. Interphase monopolization by a service needs to be monitored. It works on fixed broadband assignment. It is a heavily loaded method.

- ➢ **Weighted Fair Queuing - WFQ:** it is a method designed to avoid ***buffer starvation.*** It can identify predictable behavior in the delivery system. WQF provides enough priority to small flows as to avoid interruption by greater flows. In this way, time response optimization of interactive applications is accomplished. CBQ (CB—WFQ) features are added, allowing for bit reading of the IP datagram precedence, identifying in this way the required CoS and assigning more resources when higher priority is needed. It manages individual volumes per unit flow and operates through dynamic bandwidth assignation.

- ➢ **Low Latency Queuing - LLQ:** it combines **Priority Queuing - PQ** and **CB-WFQ** features, creating and managing a completely independent queue from the rest. This queue is reserved for applications with strict delay requirements such as VoIP and video (QoS multimedia). This queue is generally marked as “*absolute priority*.”

4.5. Traffic shaping.

Traffic shaping aims to control the volume of incoming traffic circulating in the network. It works at a network access point and is capable of controlling the individual transmission speed of each flow. It uses two methods:

- ✓ Leaky - bucket
- ✓ Token - bucket

The **leaky – bucket method** was mentioned previously and transforms a sum of incoming traffic of variable speed (gusts) in a sum of constant and predetermined flows, ensuring internal flow predictability in the network.

All incoming traffic exceeding the sender performance capability of the device is dropped (erratic gusts). It is not an appropriate method to handle incoming variable flows. It does not manage individual gusts effectively.

Token – bucket models, as mentioned above, manage incoming traffic distributing it among a fixed amount of tokens which represent in turn a fixed amount of Bytes. If there are no available tokens, transmission is not feasible.

4.6. Admission policy.

Admission policy techniques provide access control by traffic discrimination. It is based upon the contractual bases of the service. Incoming traffic filtering is based on used protocol type (UDP, TCP, IP) or according to IP addresses or particular ports, among others.

It is a primitive method, with little selectivity. It only grants or denies access to the network. Admission control and **traffic shaping** should operate simultaneously and in a coordinated way. This is essential to stabilize network parameters and to maintain the traffic generator bound to, not only its average values, but also its gust indicators.

Admission control can also deny access when the network resource reserve required compromises network stability. Previous to granting access, the system checks the resource availability along the path involved in the connection.

Packet acceptance or denial will depend not only on the availability of resources, but also on the need that the required level of service shall not exceed the contractual limits established between the client and the network service provider.

4.7. Severe Congestion Control.

With **severe congestion control** techniques, it is possible to apply some of the following two methods before the network risks congestion collapse.

- ✓ ***Random Early Detection – RED***. Its main objective is the detection of congestion collapse. The mechanism consists of constant monitoring of queue lengths at traffic delivery. When the length of one queue exceeds the security boundary established, RED makes a random selection of packets and drops them.

  An important feature is that if the deletion is made over TCP traffic, the endpoints will automatically reduce their transference rate when detecting the missing packets, decreasing in turn incoming flow. If the dropped packets are not TCP, only immediate collapse is avoided, but the main causes are not addressed. This method is considered fair because flows with greater amount of packets loading the network are more likely to be dropped.

- ✓ ***Weighted Random Early Detection*** **- WRED** drops packets according to a fixed criteria. Flows will be selected for deletion upon access.

  The probability for deletion can be varied in this case in function to the precedence bits. Traffic packets with less priority will be discarded first. As in the RED case, it will be effective if TCP flows are drop.

5. ARCHITECTURE AND DIFFSERV IMPLEMENTATION- RFC2475.

5.1. Classification.

Under this architecture, classification is possible by granting a particular differentiated *state or category* to a certain data flow using the IP datagram DS field. Packets are classified and marked at the network limit nodes or by an access host and therefore receive a particular treatment when sent to the following node.

The classification category is maintained along the path, from endpoint to endpoint. It is complemented with admission policies and shaping techniques. All these elements, along with the ***per-hop behavior – PHB*** technique applied to properly labeled packets, constitute Diffserv.

The system must have enough granularities to be capable of applying individualized treatment to multiple flows competing for common network resources, such as bandwidth or device buffers. The value labeled at the DS field, IPv4 or IPv6 is used to associate a definite behavior executed at the node. It should be noted that **Diffserv** works unidirectionally. Thus, symmetric flows are to be declared explicitly.

5.2. Diffserv Components.

They are:

- ✓ **Per-Hop-Behavior - PHB:** Node behavior towards determined flow profiles.
- ✓ **Traffic Conditioning Agreement - TCA:** agreement on traffic conditions by which classification rules are specified considering different determined traffic profiles. These rules are applied to selected packets by the classifier.
- ✓ **Service Level Agreement - SLA:** Service contract between a user and a service provider compatible with Diffserv. SLA can subscribe to all or part of a TCA

- ✓ **Dominion DS:** Set of nodes operating under the same PHB definition, granting the same service features and sharing the same policies.
- ✓ **DS ingress node:** access node to de DS Dominion, responsible for having incoming traffic to the dominion match some particular TCA (Edge Node).
- ✓ **DS egress node:** Exit node from the DS Dominion, in charge of delivering a flow of packets to the following dominion shaped with a compatible TCA of the following dominion (Edge Node).
- ✓ **DS interior node:** node belonging to a DS dominion and that is neither ingress nor an egress node. An edge node may be an interior node in some of its interphases.
- ✓ **DS region:** DS adjacent set capable of offering Diffserv.
- ✓ **Boundary link:** Data link between edge nodes from two different DS dominions.
- ✓ **Micro flow:** Individualized flow between applications and identified by ports, origin address, and corresponding destination, in addition to the network protocol.
- ✓ **DS Field:** TOS octet in IPv4 and Class traffic Octet in IPv6. In both cases, the mentioned fields contain a DS Code Point.
- ✓ **DS Code Point - DSCP**: Part of the DS field used to select a PHB.
- ✓ **Pre-mark:** to place the DS code point in a packet before entering the DS domain.
- ✓ **Re-mark**: to change the DS code point in a packet placed previously with a marker according to a TCA
- ✓ **Classifier:** Packet classifier according to content in one or some of the network protocol fields; such as: ports, origin and destination address, ID protocol, and others.
- ✓ **Marker:** Device used to place a Code Point to the packets sent by the Classifier according to specific rules for each type of flow.
- ✓ **Shaper:** Device used to generate controlled delays at a particular flow to make it follow a specific traffic profile.
- ✓ **Dropper:** Packet discarder following pre-established policies.

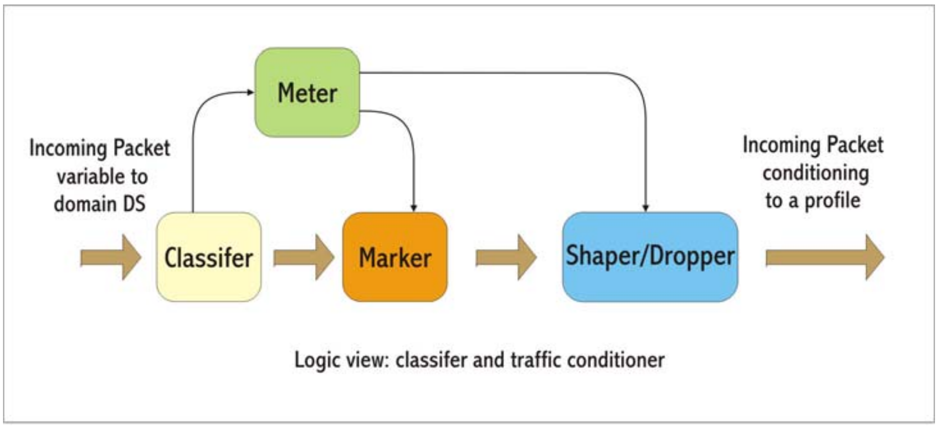

Figure N° 5. Network procedure for traffic conditioner

- ✓ **Meter:** Device used to handle variable measuring processes, such as speed or some other temporal phenomena for multiple purposes. Marker, shaper, and dropper operation rely on this instantaneously measured value.
- ✓ **Traffic Conditioner:** it conditions traffic. It can contain meters, markers, droppers, and shapers. Placed generally at the edge nodes.

A traffic conditioner diagram is depicted in Figure N° 5. It can re-mark or drop packets. In this way, the features of the flow are altered temporarily following a particular traffic profile, observing in this way the so called ***Traffic Conditioning Agreement***.

5.3. Operation of the Differentiated Services Model.

Network incoming traffic is ordered with a special device called *Classifier*, and possibly taken by a *Traffic Conditioner*, becoming associated to a pre-established traffic profile. Each profile, called *Behavior Aggregate,* is identified with a unique *DS code point.*

Once at the network core, the packets are forwarded according to per-hop behavior related to a *DS code point.* Finally, Diffserv is accomplished sending the packets following a combination of *Traffic Conditioning* and forwarding them with *Based Forwarding* - PHB

Edge nodes as well as interior nodes must be capable of applying the correct PHB to each packet based on the corresponding packet DS code point. Otherwise, network behavior becomes unpredictable.

Diffserv pervades beyond a particular DS domain, establishing Service Level Agreements – SLAs – between the previous and following dominions in relation to the one under consideration. TCAs between DS dominions derive from the SLAs established among them.

6. CONCLUSIONS.

These types of networks offer significant advantages over other alternatives:

- ✓ **Voice, data, video, critical and transactional mission applications transport capabilities**, with the required bandwidth and security level for each department in the company or the organization. Therefore, management and operational capabilities of the company increase while concentrating and controlling all communication services in a single corporate network. Resources are optimized as well.
- ✓ **Class of Service incorporation**. Allowing for effective bandwidth management. Improving real time communications[1]and avoiding session failures in applications.
- ✓ Providing VPNs in an IP network framework. Guaranteeing that all departments shall have **a private IP number** completely isolated from the Internet and other IP networks from other organizations connected to the same public transport network. In this way, the company obtains **an appropriate security level,** without requiring Firewalls or IPSEC like VPN implementation does over internet.
- ✓ Because it is an IP network[2], **it is independent from access media**, compatible with last mile technologies as varied as: Ethernet, SDH, digital point to point, ADSL, internet encrypted tunnels, telephone accesses, among others. This feature ensures long term technical continuity while protecting investments.
- ✓ By prioritizing traffic, **over-dimensioning the network becomes unnecessary.** In case of saturation, file traffic and applications are prioritized over the internet.
- ✓ It provides **accessing to the corporate network** from any place with internet access.
- ✓ It maintains an adequate bandwidth for each application and a **minimum delay** for critical missions, even during peak periods.
- ✓ Capability to provide **different classes of service.** Ensuring the following traffic types:

[1] Such as VoIP, and Videoconference.
[2] i.e., a Layer 3 service.

- Multimedia traffic (real time QoS): Minimum delay and jitter. Delayed packets are dropped.
- Videoconference Traffic (real time QoS): Similar to multimedia, but with less priority.
- Priority Data Traffic (Remain QoS): Traffic sensitive to Time Out, for software which might require it.
- Normal Data Traffic (Remain QoS): File transfer traffic, databases and systems in general.
- Low Priority Traffic (No QoS): E-mail and internet.

7. ACKNOWLEDGEMENTS.

The financial support provided by *Agencia Nacional para la Promoción Científica y Tecnológica* and *CITEFA* (Project PICTO 11-18621, Préstamo BID 1728 OC-AR) is gratefully acknowledged.

8. REFERENCES.

[1] Fusario R., Carrara E., Mon J., Castro Lechtaler A. y García Garino C.: An Overview of MPLS Technology: Quality of Service and Traffic Engineering 23-34, II Workshop de Arquitecturas, Redes y Sistemas Operativos. Proceedings of CACIC 2007, XIII Argentine Congress of Computer Science, Corrientes, Universidad del Nordeste, 2007, ISBN 978-950-656-109-3
[2] Alvarez, S. MPLS TE Technology Overview chapter 2 in QoS for IP/MPLS Networks. Cisco Press. 2006.
[3] Canalis, M.S. MPLS "Multiprotocol Label Switching. Una tecnología de backbone para el siglo XXI. Jornadas de Informática del Noroeste Argentino. JINEA 2002. UNNE. 2002.
[4] Pepelnjak, I., Guichard, J. MPLS and VPN Architectures. Cisco Press. 2001.
[5] Rosen E. Multiprotocol Label Switching Architecture. RFC 3031. January 2001.
[6] Stallings, W. MPLS. The Internet Protocol Journal, 2-14, Vol. 4, N. 3, 2001. Available at www.cisco.com/ijp.
[7] Awduche D. Requirements for Traffic Engineering over MPLS. RFC 2702. September 1999.
[8] Xiao X. Traffic Engineering with MPLS in the Internet Network, 28-33, Vol. 14, N. 2, Mar/Apr 2000. IEEE, Doi 10.1109/65.826369.
[9] Blake, S.; Black, D.; Carlson, M.; Davies, E.; Wang, Z.; Weiss, W. An Architecture for Differentiated Services. RFC 2475. 1998.
[10] Turner, J. S. New directions in communications (or Which Way in the Information Age). IEEE Communications Magazine, 24, 8-15. 1986.