A comparative study of the congestion avoidance mechanisms in modern communication nodes

C. MITROI
str. Segarcea nr. 3, bl. A12, sc. E, etj. 1, ap. 64, sector 6, Bucuresti

Todays communication networks are based on modern technologies. In order to attend highest performances the nettwork's infrastructure is composed mainly on reliable support like optical fiber. The nodes within networks are also based on optical intefaces. Even so due to increased service requirements it must apply some differentiation mechanisms between different service classes. Whether we talk about the Internet network, which has exponentially evolved or about an organization's network, it is obviosly that quality of service (QoS) assurance is a great challenge for the network's administrators or communication managers. This paper aims to present in a comparative manner, the mechanisms involved in QoS assurance within network's elements in the moment that congestion is already present.

(Received August 25, 2011; accepted October 20, 2011)

Keywords: Optical networks, QoS, Congestion avoidance

1. Introduction

In the present, modern communication networks are based on heterogenous resources, which includes a great variety of protocols used by different application. In order to fulfill user's service requirements the networks must deal with reliable resources. The advantage of optical infrastructure is obviously, fiber optics parameters helping networks to achieve high transmission speeds [1].

Even so it could be some moments when because of higher service requirements the nework could experience congestion. When the network is planned to support a great variety of traffic and some kind of transport capacity limitation between network's nodes exist, it must be compulsory taken into account the possibility of node's congestion appearence and accordingly developped some kind of mechanisms who is responsible for the right treatment of each traffic type. Only so it could be assured quality of service to the customers.

2. Importance of the QoS assurance mechanisms în the network's nodes

The aim of the QoS assurance is an end-to-end approach of that kind of performance parameters which define required user's expectance. The specific end-to-end architecture is illustrated in *figure 1*, the generically entity named service user can be a terminal, a server or a human operator and the transport medium could be represented by a single communication provider or a sum of such providers. Note that the infrastructure involved in services transport it shall be assumed as an optical transport medium.

As it can see inside the architecture coexist two plans, the management and control plan, which is created in order to create and respect some rules and policies defining the access and usage mode of communication resources, and the operational plan, which is designated to the enforcement of these principles and rules into the network's elements.

Inside the operational plan it concerns about the QoS mechanisms, which are specific to each network element and who is responsible to data flows treatment. From this point of view the principal mission of the Qos mechanisms is to assure required bandwidth, latency and jitter to reffered service. The most important mechanism's categories are related to congestion management and avoidance, and also traffic modelling between network's nodes.

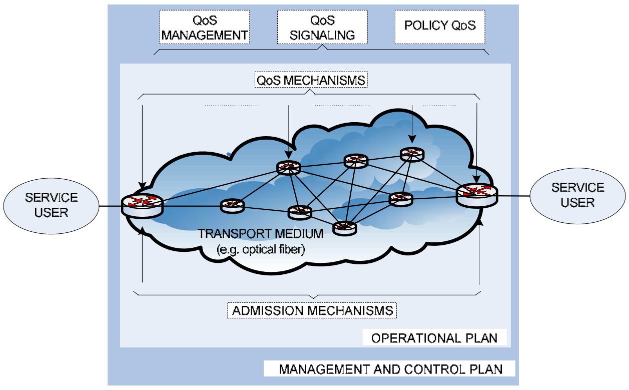

Fig. 1. End-to-end QoS architecture

Congestion management or the queuing discipline represent the form in which an network element has the ability to treat a traffic overflow, which normaly arrives on the node's incoming interface, using an algorithm in charge of data flows splitting and placing into waiting queues and priority methods defining on the element's output interface for these flows.

Congestion avoidance represent the network element ability to monitor the data flows traffic and to controlled discard of some specific flows fragments (packets) in order to prevent the congestion appearance.

Traffic modelling (traffic policing and shaping) refers to a traffic conditioning, which is used in communication nodes in order to controll traffic rates. Rates are compared with reference values, configured in network elements and then some mechanisms are activated in order to conform the traffic to these reference values.

As regarding end-to-end delay, this encompass a sum of latencies, which are produced by communication network elements [2], according *figure 2*.

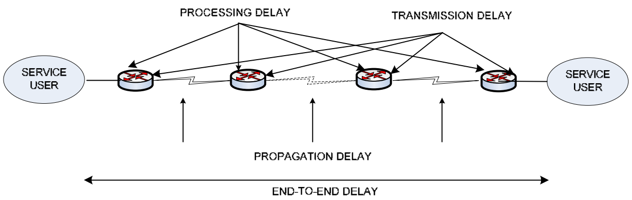

Fig. 2. End-to-end delay components.

Delay classification and their influence to the total end-to-end delay are ilustrated in *table 1*.

Table 1.

Delay classification		Appearance place	Impact to total end-to-end delay
Processing delay		Node input interface	Given by data packet switching speed between input and output interface according to some informations reading, which are inside the packet (packet destination, particular treating mode of packet based on markers like e.g. DSCP and EXP field. His value depends in a grat measure on processing speed of various node command elements (processor, memory).
Transmission delay	waiting	Waiting queue	Given by queuing time within waiting queue. His value depends on congestion management and avoidance mechanism implementation and also on node's egress interface bandwidth.
	serialization	Node output queue	Given by forwarding speed outside from output interface (bit transmit speed outside from node). His value is inversely proportional with egress interface bandwidth.
Propagation delay		Internodal connexion	Given by propagation speed in the transmission medium. Generally speaking has a constant value, except satelit connexions, case when because the distance his value is considerably greater.

3. Congestion control mechanisms

Congestion appears in a network node in the moment when a packet arrives at node's output interface much more rapidly that he can be transmitted. That output interface is congestioned if an arrived packet must wait the transmission ending of a precedent packet.

The congestion related delay can vary from the time interval necessary to finish the last bit transmission belonging to the precedent packet, to infinite, when packet is discarded, because the waiting queue is already full. Congestion maximisation within a network is important through the influence that the packet treating speed, respectively the packet discard probability has to the latency and jitter between sender and receiver.

Generally speaking it exists two classes of mechanisms which are used in order to control the congestion between network's nodes [3].

Waiting queue management (discipline), which is used in order to control the bandwidth value of an output port according to each class of service, in other words service class control in case of a limited bandwidth;

Congestion avoidance management, which control the packet number within a waiting queue (queue depth) through establishment of the moment and the packet type, which must eliminated, in other words service class control in case of a limited queue buffer length.

Even if the two mechanisms are interdependent, they use different concepts. While the first allow congestion management through control of the bandwidth allocation control into the output interface between different class of services, the second prevents congestion through control which is performed on the mean value of the waiting queue length (*figure 3*).

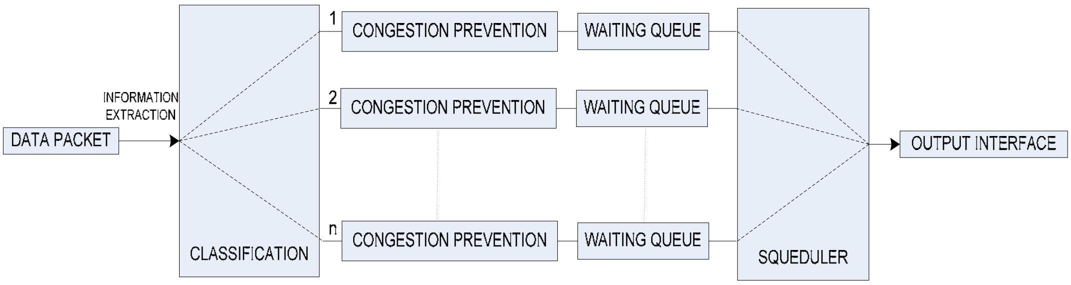

Fig. 3. Operation mode of congestion control mechanisms.

Below it be presented concisely the main targets which are followed by congestion avoidance mechanisms, the involved algorithms and their function mode, and also a comparation between the efficiency regarding service performance parameters assurance (delay, jitter, packet loss).

4. Congestion avoidance mechanisms

This kind of mechanisms deals in principal with congestion anticipation and avoidance strategies. Because of the finit length of waiting queues, these could be filled up and it could appear situation when those capacity are exceeded. În the moment of queue's filling next arriving packets will be discarded.

The congestion avoidance mechanisms must action on one hand to queue's filling aviodance so that higher priority packets could be forwarded within queues, and on the other hand to create rules in order to packet discard in case of queue's filling, with respect to the packet priority level. In a network with bursty packet flow is necessary that queues filling level converge to zero in order to absorb these bursts without packet discard.

Choosing a mechanism must be performed according to latency level, which is imposed by waiting queue. Higher queue lengths assure less packet discard, but these will introduce higher latency levels. Opposite, lower queue length will produce the reverse effect.

The algorithms used within this mechanisms cathegory are [4]:

a) *tail drop*;
b) *random early detection* – RED;
c) *weighted random early detection* – WRED;
d) *explicit congestion notification* – ECN;

4.1. Tail drop algorithm

Tail drop algorithm treat equally whole traffic, without any distinction between service classes. In the moment when new packets arrive in a already filled queue these are discarded untill queue will be free.

The main advantage of this algorithm is the easiness in his implementation on one hand and on the other hand the number of discarded packets will be reduced when queue's length is long enough. Anyway, an excessive queue's length could drive to latency growth of the forwarded packets.

The mechanism's limit is done by his incapacity to absorb bursty traffic, because due to the fact that packets are not discarded untill queue capacity is 100% filled, a succesive burst could not be processed within the queue. This situation could drive to the effect that a low flows number coud monopolise whole queue capacity, blocking another flows to access the queue.

Tail drop algorithm is not recomanded for TCP-based traffic, because this protocol is packet drop sensitive. Case when TCP source is notified about congestion start the sender will reduce automaticaly his transmit speed. When many concurent TCP sessions are forwarded in a congested queue, the algorithm will impose that all sessions have to reduce simultaneous their transmit speed., driving to global synchronisation phenomenon. This produces criticla traffic oscilation that conduct to an inefficient bandwidth use on egress port interface.

4.2. Random early detection – RED

The random early detection algorithm [5] was proposed in order to improve tail drop algorithm. The scope was to treat congestion in a manner rather preventive than reactive. RED assures these requirements through traffic load monitoring within queues and packet discard using stochastic algorithms. RED is based on the fact that majority traffic uses transport level implementation which are sensitive to packet loss. An example is the TCP protocol, which rapidely answer in the moment of packet discard through transmission speed reduction. The mechanism's scope is to control mean queue length and signalling to sender and receiver that it is neccessary a transmission speed reduction. Packet loss is time distributed under condition of maintaining a low packet level within queue, conducting so to a peak traffic absorbtion.

The advantage of RED mechanism towards tail drop is that packet dropping could begin when a configurable congestion level is set, the dropping probability beeing detrmined by three factors: the low threshold, the high threshold and the discard probability denominator (*figure 4*).

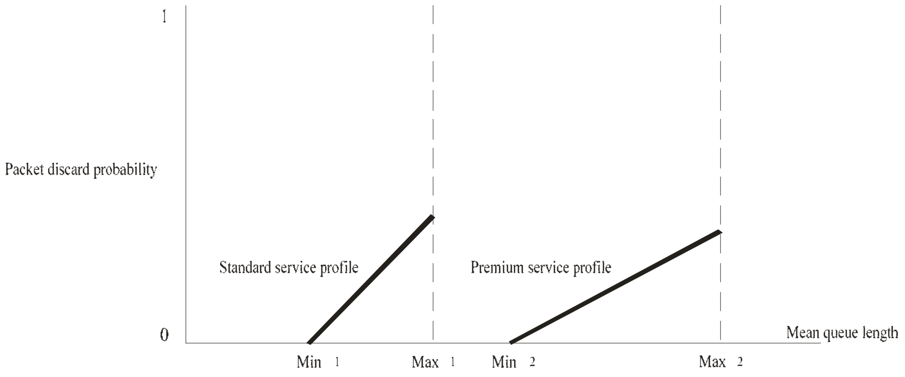

Fig. 4: Discard probability in RED mechanism for two service level

When queue packet number exceeds minimum threshold, RED begins to discard packets. The discard speed liniary increase in proportion as queue length approach to maximum threshold. The probability indicator that shows when a packet could be marked in order to be discarded represents packet percent which are discarded in the moment of maximum threshold achieving. For queue length levels which exceeded maximum threshold, RED discard all arriving packets.

The minimum threshold must be choosen in order to have a maximisation of communication channel. If the minimum threshold is too low packets could be unnecessary droped driving to an inefficient use of communication channel. Another important aspect is linked to the difference between the two thresholds, that must be enough great so that global synchronisation must be avoided.

The main RED advantage is that he idetifies the incipient congestion phases and answer through random packet discard. Also because the mechanism doesn't wait untill the queue is totally full, the queues can accept bursts of traffic.

Packet discard is done in an equilibrate manner without need to memorise the parameters of each flow. As an example, in case of a 20% drop probability configuration when queue is filled 50%, a flow which transmit 40% of his packets iin a queue will be more affected as a flow which transmit 5% of packet.

RED limits manifest concerning non TCP flows, which doesn't modify those transmit speed even if RED will drop some of packets. In this case is more efficient the use of tail drop mechanism. Also mechanism complexity could be rised in order to reach expected performances, all the more so as a wrong configuration will drive to an inefficient bandwidth use. It exists many research works which have developped many variation of RED mechanism.

Adaptive RED – ARED [6] assume a more or less aggresive behavior of originally RED through discard probability according to mean queue length. The aquired advantage lies on automaticaly configuration of parameters according to traffic levels modification. (*figure 5a*).

Stabilized RED – SRED [7] assume a drop probabilitywhich depend on active connexions and on instantaneous queue length.

Dynamic RED – DRED [8] which propose to mantain queue length about a limit level which is defined by user.

G(ently)RED [9] introduced in order to reduce the unwanted queue oscillations through usage of a subunitary value of discard probability between the two thresholds and of an unitary value at the moment when queue length is equal with the double value of maximum threshold (*figure 5b*);

BLUE [10] which uses for congestion meassure the degree of link usage level and packet loss instead of queue length value.

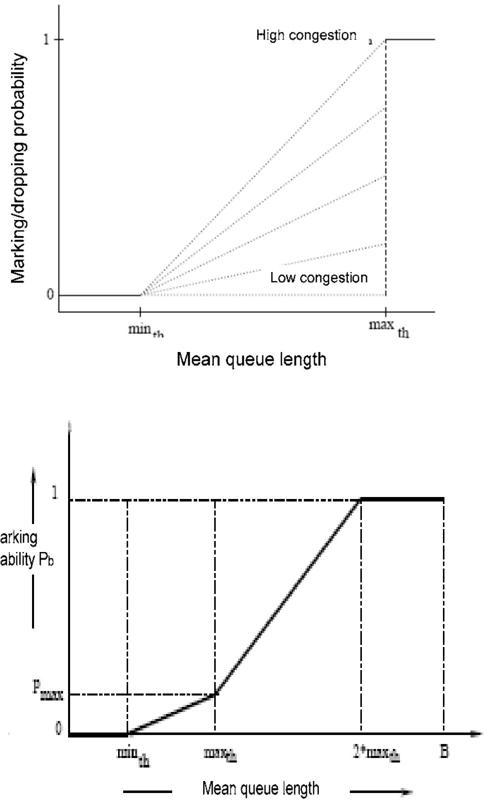

Fig. 5. Two variants of RED algorithm (ARED – 5a [7] and GRED – 5b [10]).

4.3. Weighted random early detection – WRED

A RED mechanism extension which use his capabillities an introduce also a weight linked to some specific packet characteristics is WRED algorithm, who assure in that way a preferentialy treatment of high priority packets (*figure 6*).

WRED mechanism is used mainly for TCP compatible flows, which assure the retransmission when the packets are dropped. An advantage of WRED mechanism against RED mechanism is that avoiding simultaneous dropping of higer number of packets belonging to different flows WRED protect flows against global synchronisation phenomenom appearance, conducting to a high degree of link utilisation.

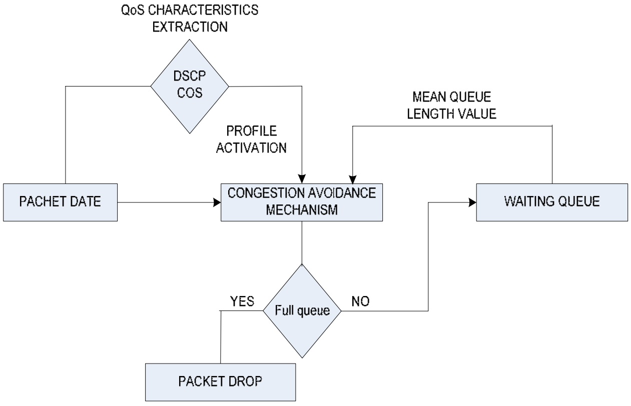

Fig. 6. WRED mechanism

In case of an transmission medium which use the IntServ model, WRED mechanism will drop first the packets that are IntServ non conform. For DiffServ model selection will be made based on DSCP field, this been the motivation wherefore the mechanism generally applies to nodes which belong to network's core.

RED and WRED mechanisms are usefull in case of a TCP traffic. For non TCP flows the proper mechanism is *flow RED* which classifies ingress traffic in flows depending of parameters like source and destination adress, ports and controls traffic according to packet number belonging to a specific flow, which reside within egress resources. Through this it could be determinate which flows will monopolise the resources and also it could be taken measures in order to reduce bandwidth consumption.

4.4. Explicit congestion notification – ECN

This mechanism try a different approach towards RED, which answer to congestion appearance through packet discard. ECN answer to congestion appearance through marking of discard predisposed packets ECN and marker transmission to the destination, which will notify the source in order to slow transmission rate. [11].

ECN suppose the definition of two bits in IP field, namely bit 7 and 8 within Differentiated Services Code Point (DSCP) field. Through combination between bit 7 (*ECN-Capable Transport* – ECT) and bit 8 (*Congestion Experienced* – CE) it will be obtained three situation: non-ECN capability (00), ECN capability (01 sau 10) and congestion appearance (11). To protect the network against nonconform on insensible TCP flows the algorithm will eliminate anyway, at the moment when the maximum threshold is excedeed, the packets instead of CE bit setting.

At the moment when an ECN indication is received by the source, this will answer with congestion window decreasing. Concerning the transport level ECN mechanism requires also an improvement in order to determine ECN capabilities for all ending points for further CE bit setting.

Three supplementary configurations are necessary in the TCP protocol header in order to facilitate the ECN mechanism implementation. First of all refers to ECN compatibility indication to the parts which are linked, this indication being activated whithin connexion activation. The second is represented of an ECN-Echo flag definition within TCP header so that when the destination receives a CE bit configured packet he will informate the source about CE bit receiving. The last configuration reffers to a congestion window reduced flag – CWR definition so that the source who transmited a CE bit further marked packet shall inform the receiver that his transmission window was half reduced.

Supplementary towards standard information exchange within TCP connexion initialisation the sender and the receiver exchange also information about their ECN-capability. If source transmit a TCP SYN packet who has both flags ECN-ECHO and CWR set, the destination will be notified of the bi-sense source participation within ECN mechanism. If the destination answer to previous packet with a TCP SYN-ACK packet with ECN-Echo flag set and CWR flag unset, the source will be notified with destination ECN capability. *Figure 7* illustrate the ECN negociation process phases between source and destination.

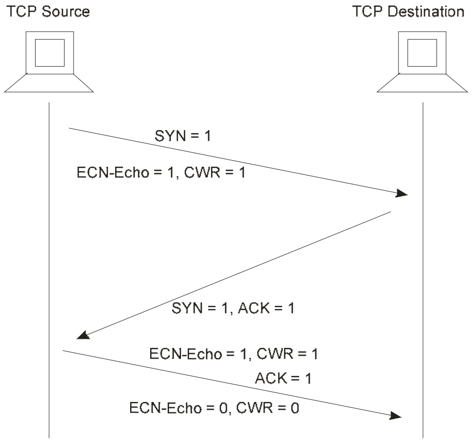

Fig. 7. The ECN negociation phases within TCP connexion initialisation

At the moment of packet transmission, the source sets ECT bit within IP header to indicate towards network nodes that they could mark the packet through CE bit setting, when congestion appears. When the destination receives a packet with both ECT and CE bits set this will set also ECN-Echo flag within TCP header of the ACK answer packet. To prevent packet loss or discard, the destination transmit ACK message untill she receives a CWR set flag packet.

The ECN mechanism advantages are similar with RED mechanism in order to eliminate the TCP global synchronisation and to assure a network bursty traffic absorbtion, additional through packet marking and not discarding it will be save transmission bandwidth till the congested node, because another flow's packets couldn't fill the already used bandwidth.

ECN mechanism limitation reffers to the modification required to compatibilise ECN with transport level.

5. Conclusion

In present communication networks a great importance to achieve a highest degree of reliable and good service is to implement quality of service. Aplications performance depend on the way of which network's administrator deals with QoS mechanisms.

In order to implement a better policy related to network congestion moments it shall consider the following aspects:

- it is important to developp an avoidance congestion mechanism, in order to deal early with the moments when the network get loaded;
- taking into consideration the various type of traffic, including real-time services, the best choice is to implement an weighted mechanism like WRED, with accurate limits for each cathegory of traffic;
- if the source and receiver are ECN-capable it's a good idea to add ECN functionality to a congestion avoidance mechanism.

References

[1] Harry J. R. Dutton, Understanding Optical Communications, IBM International Technical Support Organization, 1998.

[2] Adrian Farrel et al., Network Quality of Service – Know It All, Morgan Kaufmann Publishers, 2009.

[3] Congestion Management Overview, http://www.cisco.com/en/US/docs/ios/12_2/qos/ configuration/guide/qcfconmg.pdf;

[4] Chuck Semeria, Supporting Differentiated Service Classes: Active Queue Memory Management, Juniper Networks, 2002.

[5] S. Floyd, V. Jacobson, Random Early Detection Gateways for Congestion Avoidance, IEEE/ACM Transactions on Networking, **1**(4), Aug. 1993.

[6] S. Floyd, R. Gummadi, and S. Shenker, Adaptive RED: An Algorithm for Increasing the Robustness of RED's Active Queue Management, august 2001.

[7] T. Ott, T. Lakshman, L. Wong., SRED: Stabilized RED, in proc. IEEE INFOCOM, New York City, 1999.

[8] J. Aweya, M. Ouellette, D. Y. Montuno, A control theoretic approach to active queue management, Computer Networks, **36,** 203 (2001);

[9] S. Floyd, Recommendations on using the gentle variant of RED, Notes, 2000.

[10] W. Fen, D. Kandlur, D. Saha, K. Shin, BLUE: A New Class of Active Queue Management Algorithms, UM CSE-TR-387-99, 1999.

[11] K. Ramakrishnan, S. Floyd, D. Black, The Addition of Explicit Congestion Notification (ECN) to IP, RFC 3168, 2001;

*Corresponding author: cmitroius@yahoo.com