OPTIMIZATION OF CONGESTION MANAGEMENT MECHANISMS WITHIN CUSTOMER EQUIPMENT NODE IN AN INTRANET NETWORK

Codruț MITROI[1], Dragoş STROESCU[2]

În cazul unei rețele Intranet care utilizează şi resurse furnizate de operatorii de comunicații, apare inevitabil o limitare a lărgimii de bandă disponibile. Aceasta va conduce automat la apariția nedorită a congestiei în rețea, fenomen care poate conduce la degradarea semnificativă a calității serviciilor furnizate în Intranet. Pentru a putea asigura un nivel maxim a calității serviciilor este necesar ca în cadrul elementului de rețea Customer Equipment (CE) − granița între resursele proprii Intranetului şi rețeaua operatorului, să fie optimizate mecanismele tradiționale de management al congestiei. Articolul propune o variantă de minimizare a efectelor nedorite ale congestiei la ieşirea din Intranet pe baza simulărilor unor combinații între diferitele tipuri de astfel de mecanisme.

In the case of an Intranet network, which also uses resources supplied by communication providers, it inevitably appears a limitation regarding available bandwidth capacity. This will automatically lead to unwanted network congestion appearance, phenomenon which could produce QoS significant degradation of Intranet delivered services. In order to assure a maximal QoS level it will be necessary to optimise the traditional congestion management mechanisms within Customer Equipment (CE) node, which represent the border between Intranet's own resources and provider network. This article proposes an optimal approach which minimises the congestion's unwanted effect at the border between Intranet and provider, based on simulation of different congestion mechanisms combination.

Keywords: quality of service, congestion management, Customer Equipment (CE) node, Intranet network

1. Introduction

Because of the communication network traffic nature, it may appear the situation when data volume overflows the bandwidth aloccated for the links between network's nodes, driving to congestion appearance.

Congestion minimization in a network is very important through the way in which packet treating speed, respectively packet dropping probability within

[1] PhD student, Control and Computers Faculty, University POLITEHNICA of Bucharest, engineer, Advanced Technologies Institute, e-mail: mcodrut@dcti.ro
[2] Eng., Datanet Systems, e-mail: dragos.stroescu@datanets.ro

network nodes have an influence on QoS services parameters, especially delay, packet loss and delay variation.

In the case of an Intranet network, which depends also on a transport infrastructure belonging to a provider, the congestion associated problem is more acute, because in most cases, for economic reasons, the purchased bandwidth is less than communication services requirement. An end-to-end QoS depends in this case on the way in which a right treatment is assured for data flows associated to different services, when these flows are taken over by the transport network. It is obvious that, regardless of the transport mode over the provider's network, the congestion management mechanisms must assure a right processing of flows to the source Intranet network egress, so that these could be transparently retrieved at destination Intranet ingress.

2. Related work

Over the last period many standardisation organisms (IETF, ITU-T) proposed new QoS standards and architectures, like e.g. IntServ, DiffServ, MPLS in order to improve a diferentiated service treatment. Any of this models is based on an operational plan, in charge with the correct traffic flows treatment, which deals with specific mechanisms like queue scheduling and active queue management (AQM).

Queue scheduling is an important issue, which permit traffic prioritisation. A large variety of algorithms was proposed in many papers, from traditionally FIFO to higher complexity mechanisms like WFQ, developped in 1989 by L. Zhang, A. Demers et al. and class-based WFQ, adopted by some manufactures [1].

AQM idea is to improve TCP throughput, because TCP protocol assures network load control through a reactive approach (sending sources reduce their transmission rates only after detecting packet loss due to queue overflow, conducting meantime to important packet loss). AQM role is to detect early incipient congestion within waiting queues and to prevent queue overflow and packet loss occur. Many AQM algorithms was proposed by different authors. Beginning with the simple tail drop mechanism and continuing with well-known RED, proposed by S. Floyd and Van Jacobson in 1993, which has further different approaches, like Weighted RED, Adaptive RED [2], Gentle RED [3], Stabilized RED [4] or BLUE [5], all these algorithms try in different way to minimize congestion. Another proposal developped by IETF is Explicit Congestion Notification [6], which regards a proactive congestion signalling through a congestion early detection and further notification to sources, before queue overflow and packet loss occur.

The paper is organized as follows: starting from usual congestion management mechanisms, which are briefly described in a comparative manner in

section 2, it will be proposed, based on simulation tests performed in an environment described in section 3, a configuration mode within CE network node (section 4), which leads to a congestion minimization at the border between Intranet and provider network.

3. Congestion management mechanisms

Generally speaking, there are 2 classes of mechanisms, which are in charge of congestion control within network nodes [7], as illustrated in Fig. 1:

a) waiting queue management (discipline), which is used to control an egress interface bandwidth value according to each class of service, in other words, performing the control of service classes in the case of a limited bandwidth;

b) congestion avoidance management, which controls queue packet number (queue's depth) through an establishment of the dropping moment and the packet type which must be dropped, in other words class of services control in the case of a limited queue storage capacity.

Even if both mechanisms are interdependent, they use different basis, because queue discipline permits congestion management through the control of egress interface bandwidth allocation between service classes and congestion avoidance prevents congestion through the control, which is done on queue mean length value.

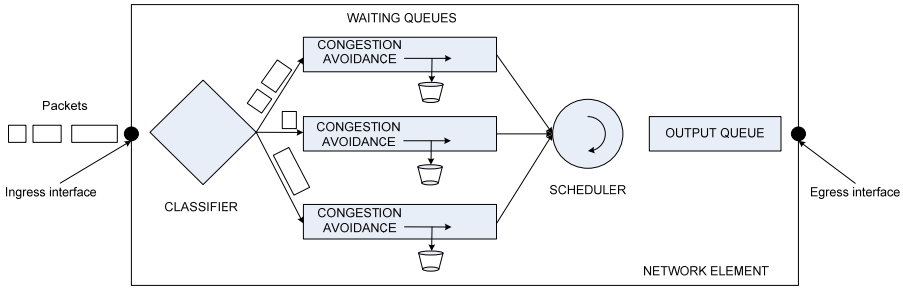

Fig. 1: Operation mode of congestion management mechanisms

Waiting queue management mechanism controls congestion through a packet classification and forwarding based on their characteristics, followed by a scheduling of packet output order.

The objectives followed by this mechanism are:

i) a balanced distribution assurance of bandwidth between different concurent classes of services. At the moment when some classes need a larger bandwidth, the mechanism must assure a specific weight to each class of services;

ii) protection assurance of different classes of services, which are associated to an egress interface, in order to block any kind of bandwidth taking-over from a less important class of services to the detriment of higher service class;
iii) assurance of an algorithm, which can be hardware or software implemented within network elements.

According to the algorithm which is used in packet treatment within waiting queues, it could be differentiated the following queue management mechanisms [8]:

a) *First In, First Out* – FIFO;
b) *Priority Queuing* – PQ;
c) *Weighted Fair Queuing* – WFQ;
d) *Class-based Weighted Fair Queuing* – CBWFQ

Congestion avoidance mechanisms put in the first place the congestion anticipation and avoidance strategies. Because queue length is not infinite, there could be many situations when the queue will be filled, at that moment any following packet being dropped.

The congestion avoidance mechanisms must act in order to prevent queue filling and, on the other hand, to create packet dropping criteria, which must take into consideration packet priority level. In a packet bursty network it is absolutely necessary that queue filling degree converges to zero, in order to absorb this kind of traffic bursts without packet dropping requirement.

Choosing between such mechanisms should be done according to delay level introduced by queue. Higher queue length will assure a less packet amount, but it also introduces higher delay values. Less queue length will have an opposite effect.

The main congestion avoidance mechanisms are the following [9]:

e) *Tail drop*;
f) *Random Early Detection* – RED;
g) *Weighted Random Early Detection* – WRED;

Below it will be briefly presented the function mode of each mechanism with their advantages and disadvantages.

a) FIFO

It is a simple algorithm and it involves a packet storage in a queue within congestion appearance and packet release in the order of their arrival. For the uncongested links this mechanism does not need supplementary configuration and could be the default mechanism [10].

The main disadvantage of FIFO is that it does not decide about packet priority, packet arrival time giving available bandwidth, switching speed and

queue assignation. In traffic peak cases, queue capacity will be totally occupied by this, other traffic categories being rejected until the queue will process supplementary packets.

Main FIFO implementations is done in hardware queues, belonging generally to output interfaces.

b) Priority Queuing

This algorithm is created for traffic which requires low delay. Then classification packets which belong to this kind of flows are put into a high priority queue, which will be treated in a preferential manner. Normally the algorithm uses a static configuration, the resulting drawback is related to a non-automatical adaptation to traffic volume modification or to weight modification between critical and non-critical traffic. If priority traffic is not constrained to respect some limits, then it will take over the resources leading to important delay for the lower priority traffic, or even to its elimination. Paradoxically, it could get into a situation, when the network serves only real time traffic [11].

c) Weighted Fair Queuing

This algorithm divides output interface available bandwidth into weigths according to different classes of services [12]. WFQ computes a final transmission time for each packet, according to transmission speed of the output interface, number of active queues that are active at a certain moment, relative weigth of each queue and length of each packet from any queue (Fig. 2).

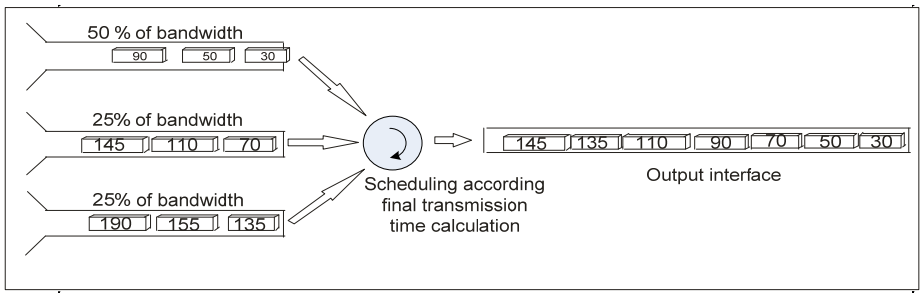

Fig. 2: Weighted Fair Queuing

The scheduler will select and transmit to the output the packet with the lowest final transmission time, which represents not the transmission time for each packet but the order in which the packets will be present at the output.

The algorithm has also some limitations, especially in connection with the implementation complexity, because of the iterative scanning requirement for all packet stage in order to calculate final transmission time.

d) Class-based Weighted Fair Queuing

This algorithm is an improvement of WFQ, which forwards packets according to a predefined classification, e.g. protocol type, access lists, interfaces, DSCP field [13]. The computing of bandwidth weight for queues could be done according to a QoS policy. It is also obvious that this mechanism could not deal alone with a great number of concurrent flows, so it must be combined with congestion avoidance mechanisms.

e) Tail drop

This congestion avoidance algorithm treats the whole traffic equally, without distinction between service classes. When new packets arrive in an already filled queue, they are dropped until the queue will be free. The main algorithm's advantage is the easiness of implementation, and the disadvantages are related especially with TCP traffic, because packet dropping in case of concurent TCP flows will lead to simultaneous transmission speed decrease of all flows, leading to global synchronisation, which produces an inefficient use of egress bandwidth.

f) Random Early Detection

In order to improve tail drop mechanisms it was proposed a RED algorithm [14], which treats network congestion in a manner rather preventive than reactive. RED monitorizes traffic load within waiting queues and drop packets through use of stochastic algorithms. RED scope is to control the mean queue length in order to absorb traffic peaks. He main advantage of RED over tail drop is that packet drop could act for a configurable level of congestion, the probability of packet dropping being determined by 3 factors: minimum threshold, maximum threshold and dropping probability denominator (Fig. 3).

At the moment when packet number within the queue exceeds minimum threshold, RED begins to drop packets in a liniar manner until the mean queue value approaches the maximum threshold. For queue length values higher than maximum threshold, RED discards all incoming packets. An important aspect is the right configuration of thresholds, because an improper diference between minimum and maximum threshold could lead to global synchronisation. With RED the waiting queues support very well the bursty traffic, because the mechanism does not wait until the queue is totally filled, so RED identifies the incipient stages of congestion and discards packets at random.

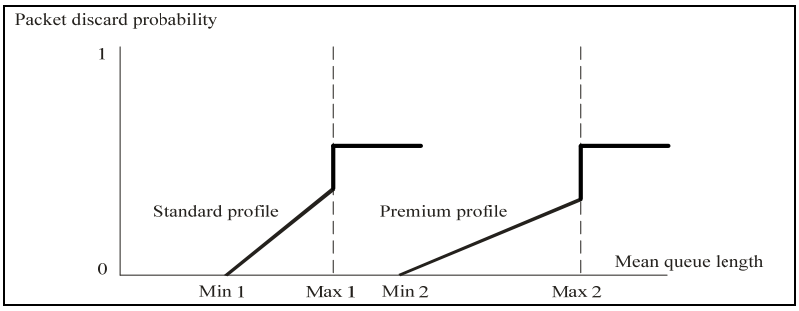

Fig. 3: RED mechanism

RED's drawbacks are related to non-TCP flows, which do not reduce their transmission speed even if packet discard occurs, in this case it is better to use tail drop.

g) Weighted Random Early Detection

A RED's extention, which uses its capabilities and introduces a supplementary weigth linked to some specific packet characteristics is WRED algorithm [15], which assures a preferential treatment to higher priority packets (Fig. 4).

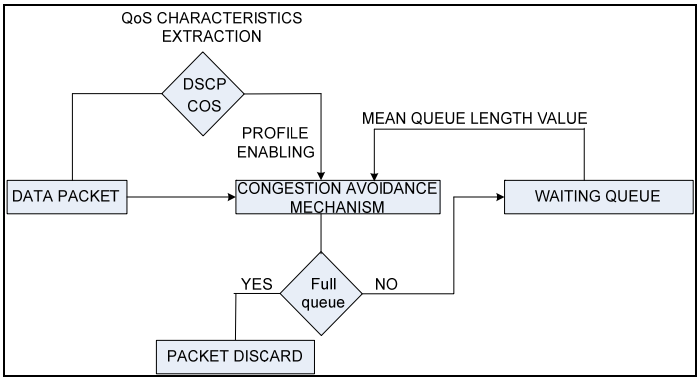

Fig. 4: WRED mechanism

WRED mechanism is mainly used for TCP flows, which assure the retransmission in the case of packet discard. An advantage of WRED compared to RED is that through an avoidance of simultaneous dropping for packets belonging to different flows, WRED assures a higher protection against global synchronisation, leading to a maximum link usage degree. For example, in the case of IntServ architecture, WRED will discard, first of all, packets which are not

conform to IntServ characteristics, and for DiffServ selection is made based on DSCP field.

As we can see, the main congestion management mechanisms have a specific role in the operational plan within network elements, which presents both advantages and drawbacks, according to traffic class, on the one hand, and the complexity of implementation, on the other hand [16]. It is obvious that the best results in congestion management are obtained through a combination of the above mechanisms, after a preliminary study about service classes.

3. Simulation environment

The simulation regards the case of an organisation which is located in two geographical sites and which uses IP telephony and videoconference real time services and also a large amount of data services in order to supply the organisation's portal and internal ERP application, divided into critical, non-critical and bulk data. The two organisations' sites are connected through a guaranteed 100 Mbps link, which is leased from a communication provider. Taking into consideration the great usage of the portal and ERP appplication, Intranet network will sustain a major traffic consisting of critical and non-critical data, without a negative influence to other kind of services. Internal bandwidth capacity is based on 1 Gbps links, which are enough to support in an optimal manner whole service areas. The only weak point is represented by the 100 Mbps link, which will be permanently congested. In this context, a number of tests were performed in order to establish the way in which different QoS mechanisms dedicated to congestion management assure the quality parameters needed for the above described organisation's services, under the condition of a round 30% overflow of the 100 Mbps link capacity, due to critical and non-critical data traffic.

Testing diagram is illustrated in Fig. 5, with the help of two Cisco 7609 routers, which play the Customer Equipment (CE) nodes role.

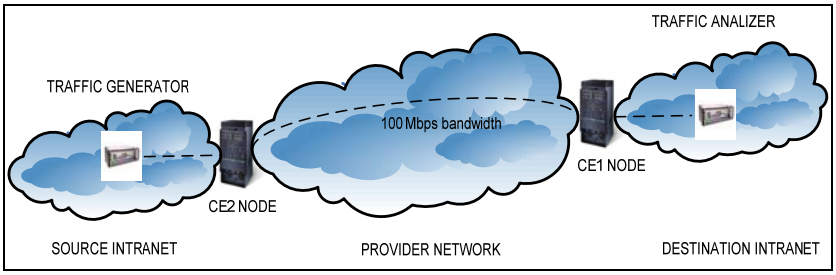

Fig. 5: Testing diagram for the congestion management mechanisms

CE 2 node represents the border between source Intranet network and provider's network and CE 1 node represents the border between provider's

network and destination Intranet network. In other words CE2 controls the Intranet's egress traffic and CE 1 is responsible with the destination Intranet ingress traffic control.

Routers are linked through a set of two 1 Gbps interfaces, whose bandwidth capacities were limited to 100 Mbps in order to simulate the guaranteed bandwidth link purchased from the communication provider.

A number of 50 different service class data flows were generated, as follows:

- 10 flows with real time service characteristics, which are marked with DSCP EF value;
- 10 flows with interactive video (videoconference) service characteristics, which are marked with DSCP AF_{41} value;
- 10 flows with critical data service characteristics, which are marked with DSCP AF_{3X} value, 4 flows marked AF_{31}, 3 flows marked AF_{32} and 3 flows marked AF_{33}, which simulate a supplementary traffic of 10 Mbps towards the initial situation, when Intranet network was designed;
- 10 flows with non-critical data service characteristics, which are marked with DSCP AF_{2X} value, 4 flows marked AF_{21}, 3 flows marked AF_{22} and 3 flows marked AF_{23}, which simulate a supplementary traffic of 20 Mbps towards the initial situation, when Intranet network was designed;
- 10 flows with bulk data service characteristics, which are marked with DSCP AF_{1X} value, 4 flows marked AF_{11}, 3 flows marked AF_{12} and 3 flows marked AF_{13};

Traffic characteristics details for the 50 flows are presented in table 1, the whole injected traffic into the CE 1 input interface being round 123 Mbps.

Table 1

	DSCP										
	EF	AF_{41}	AF_{31}	AF_{32}	AF_{33}	AF_{21}	AF_{22}	AF_{23}	AF_{11}	AF_{12}	AF_{13}
No. of flows	10	10	4	3	3	4	3	3	4	3	3
Tx rate (frame/flow)	125	250	208	278	278	278	370	370	55,5	74	74
Tx rate (Mbps/flow)	1,5	3	2,502	3,336	3,336	3,336	4,446	4,446	0,666	0,888	0,888
Supplementary traffic (Mbps/flow)	0	0	0,84	1,11	1,11	1,66	2,22	2,22	0	0	0

Within CE 2 node there were configurated the following congestion management mechanisms:

i) FIFO mechanism;
ii) WFQ mechanism;

iii) CBWFQ (class-based WFQ) mechanism with the following bandwidth distribution associated to services: voice – 15%, interactive video – 35%, critical data– 20%, non-critical data – 20% and bulk data – 10%. This kind of distribution is associated to an initial Intranet planning, which aims at real time services assurance without taking into consideration the supplementary traffic needs generated by critical and non-critical data, which appeared in the second stage;

iv) CBWFQ mechanism from point i) with the activation of LLQ (low latency queue);

v) CBWFQ mechanism from point i) with the activation of LLQ (low latency queue) and WRED congestion avoidance mechanism enabling;

Second testing stage assumed the behavior simulation of the mechanisms which are implemented on points iii), iv) and v) within a VPN-MPLS environment. Within both stages there were measured and compared QoS parameters values associated to the flows (delay, packet loss, delay variation) in order to evaluate the influence of supplementary traffic to other service categories. Traffic measure was made on CE 1 node output, which corresponds to destination Intranet ingress.

4. Results evaluation

The test data analysis conducted to the graphs illustrated in following figures, which synthetises services performance characteristics according to the implemented mechanisms.

First of all, we consider the impact of congestion management mechanisms on packet delay. As it can be seen in Fig. 6, for real time services, whether voice (DSCP EF) or interactive video (DSCP AF_{41}) delay values decrease dramatically in the case of a combination between queue scheduling and congestion avoidance mechanisms.

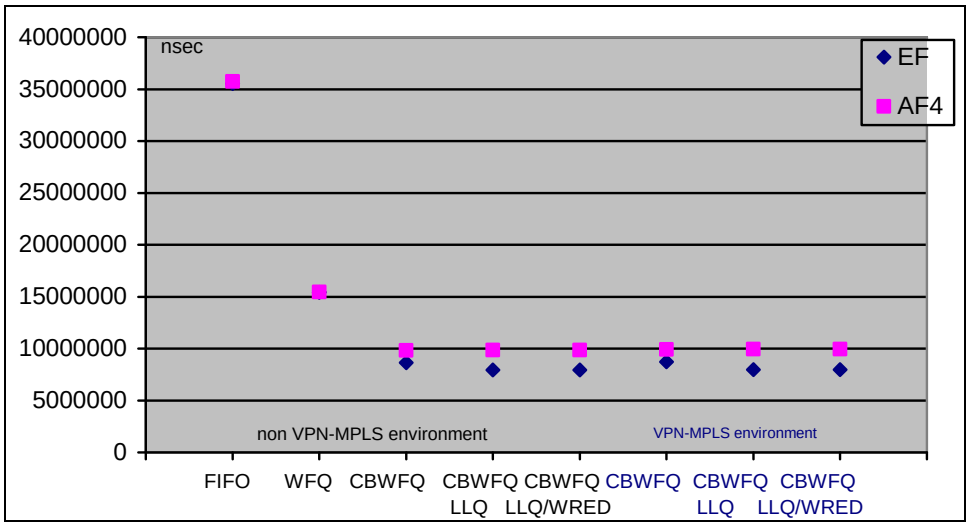

Fig. 6: Delay values for voice and interactive video services according to congestion management mechanisms implemented in CE node

The optimum solution in this case is to enable low latency queue (LLQ) for voice service within CBWFQ scheduling algorithm, the impact on delay is the approximate 3.5 times decrease towards FIFO case. The addiction of VPN-MPLS environment maintains the delay values to the same level as in CBWFQ with the obvious advantage of a higher security level assurance through services isolation within their own VPNs. In the case of interactive video service there also will be maintained low delay values using CBWFQ.

Regarding data services, Fig. 7 illustrates that in the case of FIFO and WFQ the associated delay is similar to real time services measured delay, which is a normal result taking into consideration that it does not exist any class-based discrimination.

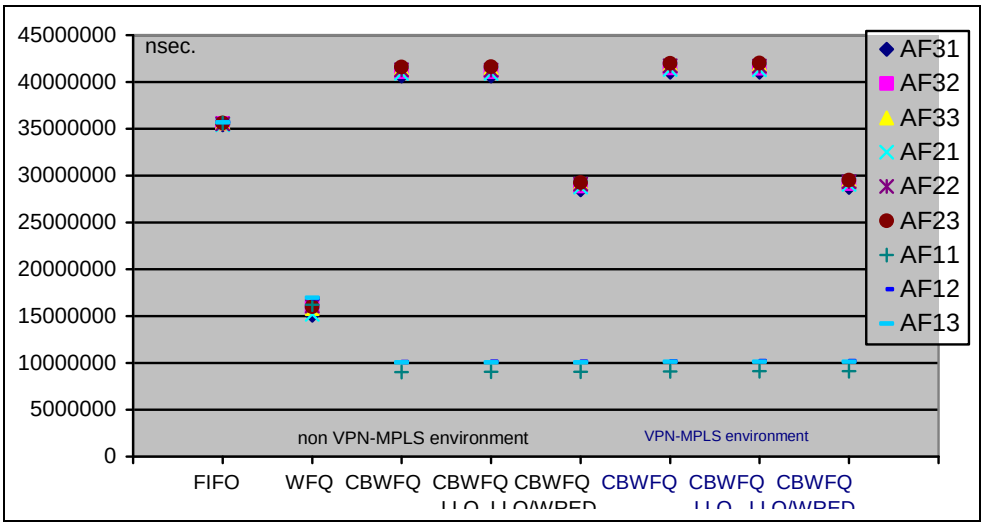

Fig. 7: Delay values for critical, non-critical and bulk data services according to congestion management mechanisms implemented in CE node

In CBWFQ implementation, AF_{1X} class maintains an approximatively constant delay level, which seems to be a normal situation due to the uncongested traffic whithin the class. The congested AF_{3X} and AF_{2X} classes have an optimum behaviour when the WRED mechanism is activated. This assures low delay values, but introduces also a packet drop within traffic classes.

The delay values' grouping in case of WRED is determined by less aggressive dropping profiles, which are implemented whithin queues for AF_{2X}, where the minimum threshold is set to 24 packets, respectively AF_{3X}, with a minimum threshold of 26 packets, both classes having a maximum threshold of 42 packets. This kind of profiles' configuration had in view the important weight, which TCP traffic has within critical and non-critical data services, a more aggressive profile driving to flows fluctuation because of global synchronisation.

Regarding the congestion management mechanisms impact on packet loss, the simulation tests illustrate a very suggestive situation, presented in Fig. 8. First of all, in the case of voice and interactive video service, the packet loss percent value is almost zero when CBWFQ or any kind of combination between this mechanism and LLQ, WRED, VPN-MPLS is implemented. As a matter of fact, FIFO or WFQ mechanism implementation is totally ill-advised for real time services, packet loss values in this cases could be over 10% for IP telephony, respectively 20% for videoconference.

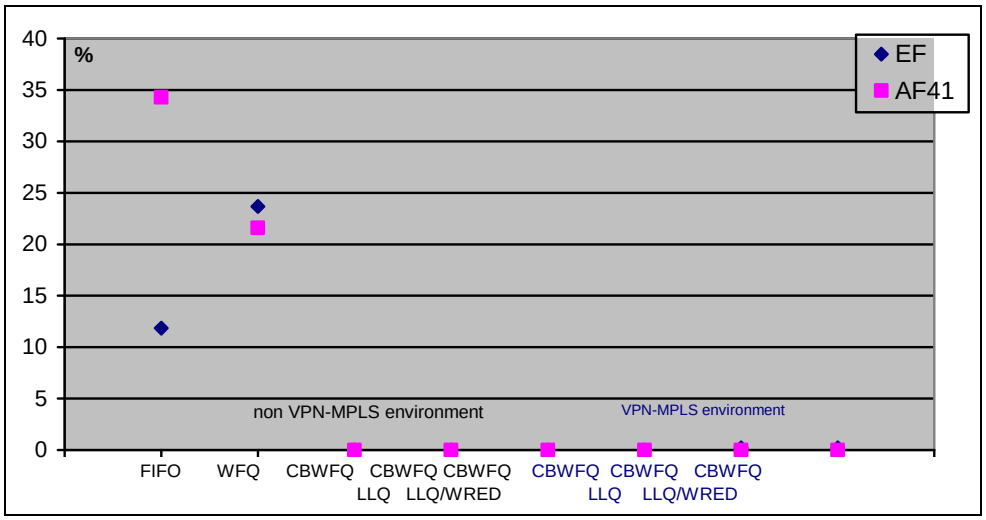

Fig. 8: Packet loss for voice and interactive video services according to congestion management mechanisms implemented in CE node

Packet loss percentual values for data services are concentrated on 3 levels in the case of FIFO and WFQ and distributed in the case of CBWFQ or a combination between CBWFQ and others mechanisms or technologies (Fig. 9).

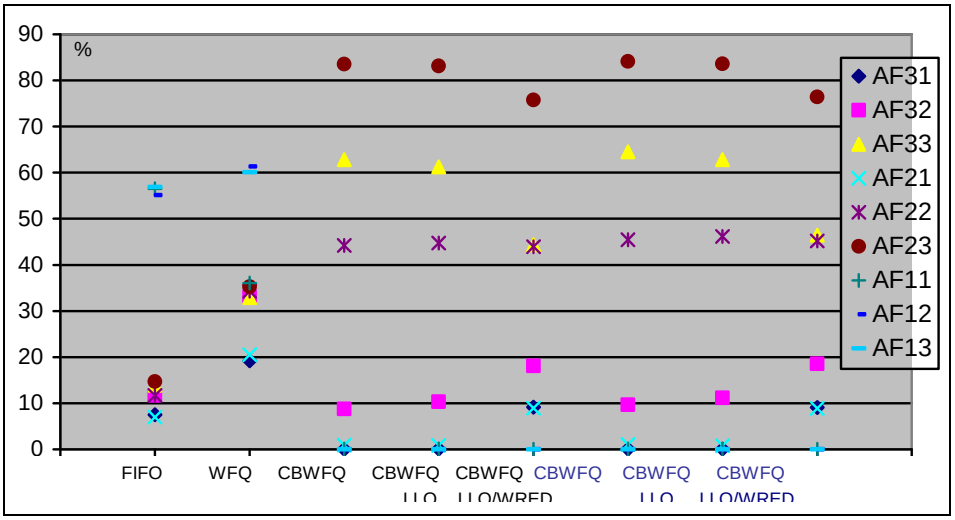

Fig. 9: Packet loss for critical, non-critical and bulk data services according to congestion management mechanisms implemented in CE node

For congestion affected services, packet dropping level increases progressively as xy indexes combination fluctuates between 31 and 23, and in WRED enabling case it will be obtained a packet dropping attenuation for inferior traffic classes (y = 3), but at disadvantage compared to superior traffic classes (y = 1, 2).

Traffic associated to AF_{1X} class is not influenced by packet loss, which is a normal result taking into consideration that this class is not affected by congestion.

Regarding the delay variation, we focus on real time service case, because this parameter is less important for data traffic. As it can be seen in Fig. 10, for real time services delay variation values are below thresholds, which are specific for this kind of services. Similar to delay, the best results are obtained through a class-based policy enabling within waiting queues.

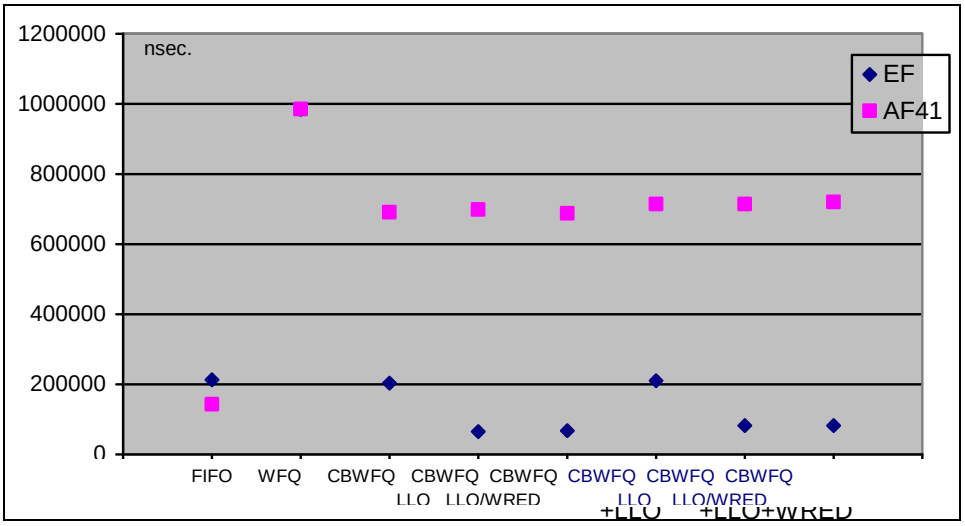

Fig. 10: Delay variation for voice and interactive video services according to congestion management mechanisms implemented in CE node

5. Conclusions

In terms of simulation results, we could conclude that QoS parameters associated to Intranet services (delay, packet loss, delay variation) are influenced by the way in which different congestion mechanisms are configured. There could be a situation when an inadequate mechanisms' configuration leads to a significant service quality degradation, especially the real time services.

The best solution is the choice of a mechanism which divides available bandwidth in fixed capacity sub-bands, according to Intranet's transported traffic

categories. In order to do this, a preliminary analysis must be made for a precise Intranet network resources' evaluation according to each service class requirement. The analysis must be followed during service delivery by a Service Level Management, which is based on an proactive architecture [17], this combination conducting to a maximum level of user's Quality of Experience (QoE).

The result is a class-based QoS architecture, which guarantees that following the Intranet network design, any kind of traffic fluctuation, whose tendencies lead to communication resources congestion, is treated within the traffic class to which it belongs, without any prejudice to other traffic categories.

Even if, on the first evaluation, the fixed sub-bands allocation might look uneconomic within some periods, when class traffic does not totally fill its sub-band, this approach, based on a solid preliminary analysis of predicted services which will be delivered in Intranet network, assures overall the maximum QoS level under limited communication resources.

R E F E R E N C E S

[1] *S. Vegesna*, IP Quality of Service, ISBN 1578701163, Cisco Press, 2001;
[2] *S. Floyd, R. Gummadi, S. Shenker*, Adaptive RED: An Algorithm for Increasing the Robustness of RED's Active Queue Management, http://www.icir.org/floyd/papers.html, 2001;
[3] *S. Floyd*, Recommendations on using the gentle variant of RED, Notes, March, 2000;
[4] *T. Ott, T. Lakshman, L. Wong*, SRED: Stabilized RED, proc. IEEE INFOCOM, 1999;
[5] *W. Fen, D. Kandlur, D. Saha, K. Shin*, BLUE: A New Class of Active Queue Management Algorithms, UM CSE-TR-387-99, 1999;
[6] *K. Ramakrishnan, S. Floyd*, A Proposal to Add Explicit Congestion Notification (ECN) to IP, RFC 2481, 1999;
[7] ***, Congestion Management Overview, http://www.cisco.com/../qos/../qcfconmg.pdf;
[8] *C. Semeria*, Supporting Differentiated Service Classes: Queue Scheduling Disciplines, Juniper Networks, 2002;
[9] *C. Semeria*, Active Queue Memory Management, Juniper Networks, 2002;
[10] *Mark Anthony Parris*, Class-Based Thresholds: Lightweight Active Router-Queue Management for Multimedia Networking, a disertation for the degree of PhD, Chapel Hill, 2001;
[11] *M. Gospodinov*, The affects of different queuing disciplines over FTP, Video and VoIP Performance, International Conference on Computer Systems and Technologies - *CompSysTech'2004*;
[12] *G. Panza, M. Grazioli, F. Sidoti*, Design and analysis of a dynamic Weighted Fair Queuing (WFQ) scheduler, International Conference on Internet Technologies and Applications, 2005;
[13] *M.J. Fischer, D. M. Bevilacqua Masi, J. F. Shortle*, Simulating the performance of a Class-based Weighted Fair Queuing system, Proceedings of the 2008 Winter Simulation Conference;
[14] *S. Floyd, V. Jacobson*, Random Early Detection Gateways for Congestion Avoidance, IEEE/ ACM Transactions on Networking, **vol. 1**, no. 4, Aug. 1993;

[15] ***, Cisco IOS Quality of Service Solutions Configuration Guide: Congestion Avoidance overview – Weighted Random Early Detection;
[16] *A. Noble*, Network Performance Technology – Whitepaper, march 2003;
[17] *C. Mitroi*, A proactive Service Level Management architecture in an Intranet network, UPB Scientific Bulletin Series C: Electrical Engineering and Computer Science, **vol. 75**, Iss. 4, 2011.