

Multi-Service Routing: a Routing Proposal for the Next Generation Internet

António Varela¹, Teresa Vazão², and Guilherme Arroz¹

¹ Instituto Superior Técnico, Portugal,
antonio.varela@tagus.ist.utl.pt,

² Inesc-ID, Portugal

Abstract. Quality of Service support plays a major role in the Next Generation Internet. QoS routing protocols must cope with service differentiation to enhance this support. This paper proposes a service aware QoS routing protocol, the Multi-Service routing, which is an extension to traditional intra-domain routing protocols. It proposes a new path selection policy that guides higher priority traffic through the shortest path and diverts lower priority traffic through longer paths when service performance degradation is foreseen. Simulations results shows that the proposed routing performs better than existing QoS routing and link-state protocols.

1 Introduction

Quality of Service (QoS) plays a major role in the Next Generation Internet (NGI), as new services and applications arise based on multimedia traffic with special requirements, demanding new service models and routing approaches [1].

The Internet Engineering Task Force (IETF) attempts to solve Internet's lack of QoS, by defining new services models. The first model proposed - Integrated Service (IntServ) provides strict QoS guarantees, but does not scale well to large networks. The Differentiated Service (DiffServ) model solved this issue and is able to assure QoS to aggregated traffic flows classified into a restricted set of service classes. Multi-Protocol Label Switching (MPLS) solution, which assures QoS support by means of traffic engineering capabilities offered below the network layer. Concerning the QoS all these technologies are expected to coexist on the NGI. Nevertheless, DiffServ will play a central role, as it offers a scalable network layer solution, being then independent of any kind of access technology or higher layer protocols.

To date, the Internet routing focuses on connectivity: routing protocols, such as the Open Shortest Path First (OSPF) or the Routing Information Protocol (RIP), are able to cope with the network impairments, but are unable to fulfill the service requirements imposed by the new kind of applications, being inadequate for the NGI. Traffic between two end points is forwarded through the same path, which is usually the shortest one, disregarding the network conditions and the QoS requirements of the associated flows. Thus, congestion arises in these

paths and service requirements can no longer be met, despite the existence of alternative underutilised paths.

Several QoS aware routing protocols have been proposed to solve these issues [2]. Should data and telecommunication networks converge around the NGI, the QoS routing problems will become very difficult to solve. First of all, this convergence leads to the existence of traffic with diverse QoS constraints in the same network and, according to [3], this may increase routing's complexity, as finding a feasible path with two independent constraints is an NP complete problem. Second, as the network state changes very often it may be difficult to gather up-to-date state information, specially in large scale environments. The use of outdated information by a routing protocol may degrade the network performance. And finally, a network where resources are shared among priority and Best Effort (BE) traffic is difficult to manage. Although performance guarantees can be assured in priority traffic, by means of resource reservation, the throughput of BE traffic will suffer, if the network capacity is under optimised, by wasting paths that may be used at least by BE traffic. Most of the QoS routing proposals are able to deal with the network state' information, but do not cope with service differentiation.

This paper aims at defining a new approach to intra-domain QoS support, where the routing protocol cooperates with DiffServ. The proposal is targeted at IPv6 networks and complaint to MPLS traffic engineering mechanisms, being particularly foreseen to the NGI.

The paper is organized as follows: section 2 presents several approaches for QoS routing; section 3 describes the routing architecture; section 4 contains the simulation results and, finally, section 5 presents the conclusions and future work.

2 QoS routing in the NGI

NGI QoS routing's support three main tasks: state maintenance, route calculation and path selection. The next sections analyse several possible approaches.

2.1 State Maintenance

State Maintenance is supported by local measurements that are performed at each node to evaluate its own state, regarding a single or multiple performance indicator. It can comprise link occupancy, residual bandwidth, delay or the availability of other resources.

A **Local State** strategy is used whenever each node only uses the information it gathers to compute the routes. Nelakuditi et al [4] uses such approach by selecting the path, that will be used to forward a flow, among a set of candidate ones, based on local information. Despite its simplicity, routing decisions are based on an inaccurate view of the network, as remote network conditions are not known.

Should local state information be disseminated through the network, a **Global State** strategy will be used. Although the network state changes very often, routing updates should be bound to reflect the longterm behaviour of the network.

Thus, instead of advertising instantaneous performance indicators, quantified metrics must be used. A simple solution was proposed within the ARPANet scope and consists in calculating the average value of the performance indicator [5]; alternatives are also used based on threshold values and hysteresis mechanisms [6] that reduce routing instability and limits the burden of traffic and processing entailed by the routing protocol.

The complexity of this Global State strategy may be compensated by the most accurate view of the network state that can be achieved when compared to the perspective attained by the Local State strategy. However, in large scale networks a less precise view of the network is accomplished, as longer delays are expected to disseminate and update the routing information. Lack of scalability also arises when the number of metrics to be advertised grows beyond a certain limit. A hybrid strategy based on **State Aggregation** can be used, where nodes are organised hierarchically into clusters; inside a cluster detailed state information is transferred, while among clusters only aggregated information circulates. Private-Network-Network-Interface (PNNI) [7] routing uses such approach, by defining a flexible hierarchical network that can grow up to 104 levels. Scalability gains leads to less optimal paths and complex routing mechanisms.

2.2 Route Calculation

Route calculation can be performed using two main techniques: source routing and distributed algorithms.

In the **Source Routing** approach each node has a global view of the network and routes are calculated at the source using this information, and piggybacked into every data packet. The entailed overhead precludes its use in large scale networks or under heavy load conditions [8].

The **Distributed Routing** attempts to solve this problem by delegating to each node the task of calculating a part of the path toward the destination. Link-state or distance vectors algorithms can be used. Their use in large networks may introduce a significant overhead, leading to the existence of hierarchical solutions, like the one presented earlier for PNNI or even OSPF.

One of the most important problems in route calculation for QoS routing protocols is related to the fact that routes can no longer be defined based on the number of hops. For instance, if the metric is bandwidth, the best route is the one that maximises bandwidth over the bottleneck link, while if the metric is delay, the best route is the one that minimises it; finally, if both metrics are considered, one needs to maximise bandwidth while reducing delay. In most of the cases the problem can be solved by using modified versions of Dijkstra's algorithms

Another issue that must be considered is the number of paths that are calculated between each pair of source and destination nodes. If a **single path** is used, routing oscillations arise, as long as multi-hop selection is used. This instability problem can be avoided by using load balancing techniques, which can be applied if **multiple paths** are calculated. In [9] it is proposed an algorithm that provides multiple paths of unequal costs to the same destination.

2.3 Path Selection

Today the Internet uses the datagram service model, where paths are selected in a **hop-by-hop** way, using the network's destination address information contained in the packet; most of the existing routing schemes are based on this principle.

Claiming that BE traffic must be routed differently than priority one, new hop-by-hop routing proposals that support service differentiation have recently arisen [10] [11]. Nevertheless, as long as the same routing tasks are performed at both edge and core network elements, a significant burden of information processing is spread across the network. In the NGI, complexity must rely on the edge of the network, in order to allow a faster processing at the core, which means that alternative path selection approaches might be more adequate.

As soon as service differentiation becomes an issue, the notion of flow is fundamental to provide QoS support and it might be used to facilitate the cooperation among routing and resource allocation policies, as a virtual service model can be envisaged [15]. By using **Flow Level** routing traffic may be easily routed according to its class of service. In [12], Nahrstedt and Chen propose a combination of routing and scheduling algorithms where priority traffic is deviated from paths congested by BE traffic. Another proposal was made in [13], where QoS traffic uses less congested paths. However, both of them use source routing paradigm, which is not adequate for NGI, as stated before. IETF has proposed a QoS routing framework [14] that performs the flow level path selection; under this proposal every incoming flow is admitted into the network, only if there are enough available resources; otherwise it is blocked. Despite the accuracy that can be achieved with this type of approach, it is very complex and may not scale well, if individual flows are considered.

Scalability may be achieved if instead of using individual the Flow Level routing, an **Aggregated Flow Level** strategy is used to perform path selection. This strategy is compliant with IPv6 standard that provides a Flow Label field in the IP packet header, and may be supported over MPLS networks. Moreover, more complex routing decisions can be rely on the edge of the network and only when traffic flows initiate their activity.

3 Multi-Service Routing

In this section the main characteristics of the **Multi-Service** routing are described. A more detailed description of its architecture can be found in [16]. In this paper a more complete study of the proposed routing protocol will be presented.

3.1 General principles

The Multi-Service routing proposal extends traditional distributed intra-domain routing protocols, by triggering routing table update cycles, whenever service fulfilment may not be accomplished due to the existing network conditions.

Smooth variant quantified metrics are used to trigger such updates, based on global network state information. To assure compatibility, standard mechanisms and messages are used in this updating process.

In spite of using an hop-by-hop approach, an aggregated flow level strategy is used, enabling a scalable and efficient solution. Aggregated traffic flows are defined at the edge of the network by assigning a Flow Label value to the respective field of IPv6's packet header. Complexity relies on the network's edge, as flow identification and maintenance are performed only at the edge routers. Unless re-routing is needed, routing decisions are taken only once, when a new flow is detected; subsequent packets are routed based on their associated aggregated flow service class.

At each time, each router may have two different routing tables: the **standard table**, describing the set of shortest paths to the destination, and the **alternative table**, describing a set of longer paths to the destination. The selection between these tables must be made according to the following set of routing policies:

- Priority traffic should be routed through a standard (shortest) path, as this one has a higher probability of assuring the required service level.
- If the network is less loaded, the remaining traffic may share the same path, as it will not interfere with the performance of higher priority traffic.
- As the network load increases, alternative paths will be found, which will be used by incoming lower priority aggregate flows, in order to meet the level of service of the already active flows and to utilize the unused network resources.
- In case severe local congestion takes place, existing lower priority aggregate flows may need to be re-routed to the alternative path.

3.2 Network State Maintenance

The Multi-Service routing was conceived to avoid complexity. Thus, it uses a Global State strategy and instead of using different measures to evaluate each node's neighbourhood state, a single and simple one was selected: the output link occupancy, which is periodically sampled. Based on the samples an indicator is evaluated using an exponentially weighed moving average (EWMA) technique.

Considering two adjacent nodes i and j and a link $l_{(i,j)}$ connecting them, a number of samples N and a weight α , the output link occupancy indicator, $L_{(i,j)}$, regarding the connection of node i toward node j , at the sampling time t_i is given by:

$$L_{(i,j)}(t_i) = \alpha * \frac{\sum_{t=t_{(i-1)}-N}^{t=t_i-1} L_{(i,j)}(t)}{N} + (1 - \alpha) * L_{(i,j)}(t_{i-1}) \quad (1)$$

Threshold values are defined and, in order to avoid nasty traffic balance oscillations effects, a hysteresis mechanisms is also considered. Whenever a threshold is reached, a quantified QoS metric is modified and the alternative routing table update procedure is triggered.

When $M_{(i,j)}(t)$ represents the value of the QoS metric between node i and j at sampling time t ; T_k represents the k^{th} threshold; H_k the associated hysteresis value and M_k the corresponding metric. At a sampling time $t_i > t$ link $l_{(i,j)}$ changes its QoS metric, as long one of the two following conditions apply:

$$L_{i,j}(t) < T_k \wedge L_{i,j}(t_i) \geq T_k \Rightarrow M_{(i,j)}(t_i) = M_k \quad (2)$$

$$L_{i,j}(t) \geq T_k \wedge L_{i,j}(t_i) < T_k - H_k \Rightarrow M_{(i,j)}(t_i) = M_0 \quad (3)$$

Two major threshold values were defined:

- **Deflection Threshold** - it acts like a type of pre-congestion alert; when it is reached, all previous traffic flows keep their paths, while the new incoming lower priority traffic flows are routed according to the new alternative routing table's paths that will surely not include the current link.
- **Critical Threshold** - it causes the removal of all low priority traffic flows that are currently crossing the critical link. This removal is done by a signaling mechanism that notifies a set of border routers to take the appropriate actions to reroute their incoming lower priority traffic flows that are crossing the saturated node's link at the time. Border routers determine new paths to those flows by deleting related ones.

Hysteresis is also defined as **Standard Thresholds**. When they are reached, it means that a steady light traffic load condition persists in the node's link and the paths containing this link will be available, again, to the new low priority traffic flows.

3.3 Route Calculation

Multi-Service routing is an extension of traditional intra-domain routing protocols, being able to use a link-state or a distance vector approach. Routing information is distributed to all routers in the domain. If a Link-State routing strategy (OSPF) is used, two independent instances of the routing protocol are executed at each node. One of them periodically transfers Link State Advertisements (LSAs), which carry the administrative metric, and updates the standard routing table, accordingly; the other one uses LSAs to disseminate QoS metric and updates the alternative routing table. In order to have multiple paths per destination, a modified version of the Dijkstra algorithm is used in each routing instance. If a Distance Vector routing protocol (RIP) is used, the same type of structure is employed: two independent instances of the protocol are used, one uses the administrative metric and computes the standard path, while the other uses the QoS metric and computes the alternative path. Multiple paths per destination for each service class leads to the utilisation of a modified version of Bellman-Ford algorithm.

Administrative information is periodically transferred to assure consistency of routing information, but also when a topological change occurs. As regarding QoS information, the network state may change very often, leading to frequent

changes in QoS metrics. To avoid a burden of routing traffic due to such situations and routing instabilities, QoS routing information is transferred periodically or when there is a change on a QoS metric that occurs after a stability period since the last change. Thus, very frequent changes are only advertised if they persist after that period of time.

Considering link $l_{(i,j)}$ and the existence of modifications on its QoS metric $M_{(i,j)}$ that occur in two instants of time, instant t_i and instant $t_i + \delta$; considering also a stability period of T ; such modification will only generate an alternative routing table update event, $Ev_{(i,j)}$, if the following condition is verified:

$$M_{(i,j)}(t_i + \delta) \neq M_{(i,j)}(t_i) \wedge \delta \geq T \Rightarrow Disseminate(Ev_{(i,j)}) \quad (4)$$

3.4 Path Selection

The Multi-Service routing path selection strategy is based on a Aggregated Flow Level strategy, being completely different from the traditional intra-domain hop-by-hop method.

At the edge of the network, each incoming new flow is classified into an **Aggregated Service Class**, according to its service class, age and ingress and egress nodes. The first packet of each flow that arrives at each node uses the routing tables (standard or alternative) to identify the next hop; subsequent packets of the same flow are associated with it at the edge of the network; their routing will be based on the flow identifier they carry and on the associated routing information, retrieved by this first packet to select the path.

Considering a packet, $pkt_{(i,t)}$, arriving at node i at instant t ; the aggregated service classes $ag_sc(z)$, where z represents a specific class and *any* a class among the existing ones; the DiffServ service classes $sc(p)$, where p represents the priority of the class (*Prio* or *BE*); the network state's conditions, from node's i perspective, $ns_{(i,s)}$, where s represents the network state (low (*L*), medium (*M*) or heavy (*H*) load conditions); the standard routing table, *Std_Rt* and the alternative routing table, *Alt_Rt*; and also the selected next hop $hop_{z,x}$, where x is the node's selected egress interface (s via the standard path and a via the alternative one), the routing policies can be defined as follows:

- $if\ pkt_{(i,t)} \notin ag_sc(any) \wedge pkt_{(i,t)} \in sc(Prio) \Rightarrow$
 $new(ag_sc(z), pkt_{(i,t)}) \leftarrow z_1; select(Std_Rt_{(i,t)}, pkt_{(i,t)}) \leftarrow hop_{(z_1, x_s)}$
- $if\ pkt_{(i,t)} \notin ag_sc(any) \wedge pkt_{(i,t)} \in sc(BE) \wedge ns_{(i,L)} \Rightarrow$
 $new(ag_sc(z), pkt_{(i,t)}) \leftarrow z_2; select(Std_Rt_{(i,t)}, pkt_{(i,t)}) \leftarrow hop_{(z_2, x_s)}$
- $if\ pkt_{(i,t)} \notin ag_sc(any) \wedge pkt_{(i,t)} \in sc(BE) \wedge ns_{(i,M)} \Rightarrow$
 $new(ag_sc(z), pkt_{(i,t)}) \leftarrow z_3; select(Alt_Rt_{(i,t)}, pkt_{(i,t)}) \leftarrow hop_{(z_3, x_a)}$
- $if\ pkt_{(i,t)} \notin ag_sc(any) \wedge pkt_{(i,t)} \in sc(BE) \wedge ns_{(i,H)} \Rightarrow$
 $new(ag_sc(z), pkt_{(i,t)}) \leftarrow z_4; select(Alt_Rt_{(i,t)}, pkt_{(i,t)}) \leftarrow hop_{(z_4, x_a)}; reroute(ag_sc_{(z_4, x_a)})$
- $if\ pkt_{(i,t)} \in ag_sc(z_i) \Rightarrow select(hop_{(z_i, x)})$

4 Simulation Studies

4.1 Simulation Scenario

The proposed routing architecture has been tested through simulations, using the Network Simulator (NS), version 2.27, which has been enhanced with additional capabilities, needed to support this new proposal. Simulations with different network load conditions were performed, using the network scenario described in figure 1 and in table 1.

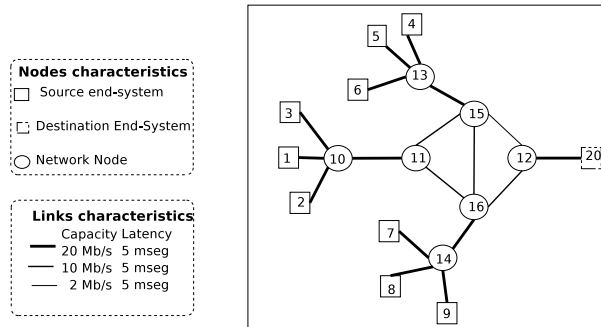


Fig. 1. Network Topology

Table 1. Traffic Characterisation

Class	Type	Number	CoS	Traffic	Src	Dst	Rate[Kb/s]	Size[B]	Total BW [Kb/s]
Prio	Single	1	EF	CBR	1	20	24	40	24
Prio	Aggregate	42	EF	CBR	4	20	24	40	1000
Non-Prio	Aggregate	[0..18]	BE	CBR	5	20	500	1500	[0..9000]

4.2 Parameterisation of Threshold Values

A set of simulations were carried out to configure the thresholds of the Multi-Service routing protocol, in order to adjust the performance of the Multi-Service routing protocol.

In the first set of simulations the Multi-Service routing supports only the critical threshold, which means that when it is reached the entire set of non-priority flows are deviated from the shortest path. This kind of situations should happen only when the network is heavy loaded and thus the threshold values tested are high (80% and 90% of the link occupancy). The threshold that offers the best performance is the one that reduces the losses and delay. As stated in figure 2, although similar results are achieved by both threshold values, fixing the critical threshold at 80% removes the transitory

spikes that happened before the path transition occurs and decreases the number of losses in non-priority traffic, which means that a more efficient network utilisation is achieved.

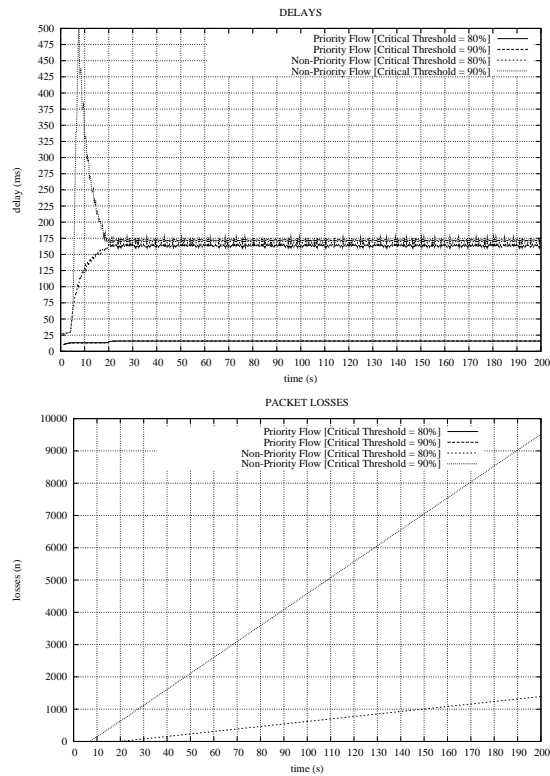


Fig. 2. Critical Threshold parameterisation

Should the critical threshold be fixed at 80%, the deflection one may be tuned. Three different values were tested (20%, 50% and 70%) and the results are shown on figure 3. If the deflection threshold is adjusted to 20% of the link capacity, incoming non-priority flows starts to be diverted too soon and longer delays are achieved for both priority and non-priority traffic. On the other hand, if the 70% value was selected BE losses will be more significant than those achieved when the deflection threshold is defined at 50%, because the modification of the paths happens too late, when the smaller capacity link (15-12) is already heavy loaded. At 50% of link capacity, both priority and non-priority traffic have a good performance, as delay is kept small and no losses occur in BE traffic.

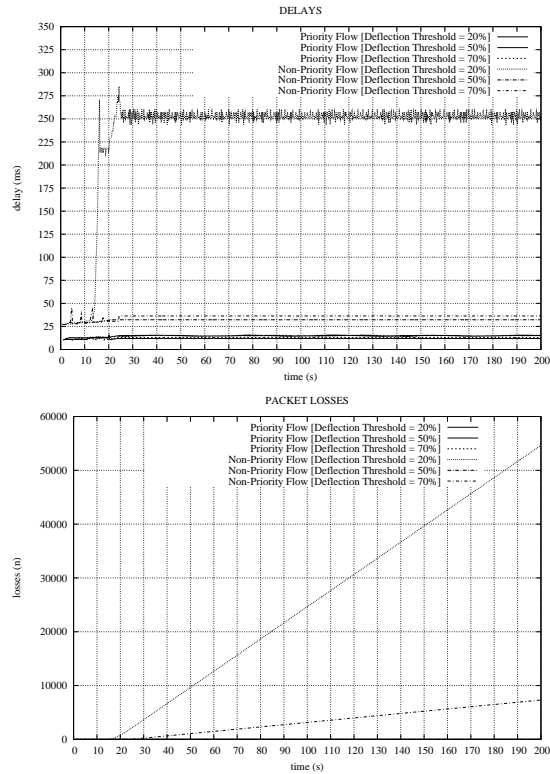


Fig. 3. Deflection Threshold parameterisation

4.3 Performance evaluation of Multi-Service routing

The performance of Multi-Service routing (MS-R) was compared to the performance offered by both the traditional link-state (LS-R) and the QoS routing (QoS-R). The results shown in table 2 illustrates the performance of those algorithms.

For both types of traffic, the Multi-Service routing is the one that presents smaller delays; throughput and losses are similar to those achieved by QoS routing, which are much better than the ones achieved by traditional link-state routing.

A more accurate view of the different behaviour of the Multi-Service and the QoS routing protocols is depicted in figure 4.

As can be stated, the Multi-Service routing also presents a more stable longterm behaviour, as no significant traffic spikes occurs. At time instant 3, the deflection threshold is crossed because the output link of node 12 towards node 15 reaches 50% of its capacity; non-priority traffic presents a slightly better performance than the one it has presented before, as new incoming non-priority flows are diverted through a longer path. As new priority traffic are still being applied to the network after that time instant, the link occupancy (15-12) stays near 80%, but only at time instant 29, it crosses the critical threshold. At this time, all the non-priority traffic is diverted to a longer path and so the critical link occupancy and the delay of priority traffic

Table 2. Priority traffic: performance evaluation

Priority traffic									
Link load	1 Mb/s			5Mb/s			10 Mb/s		
Type of routing	MS-R	QoS-R	LS-R	MS-R	QoS-R	LS-R	MS-R	QoS-R	LS-R
Delay[ms]	11.91	11.91	11.91	11.95	15.41	10807	11.98	19.9	25813.4
Losses[%]	0	0	0	0	0	80.16	0	0	95.60
Throughput[Mb/s]	0.99	0.99	0.99	0.96	0.96	0.18	0.91	0.91	0.04

Non priority traffic									
Link load	1 Mb/s			5Mb/s			10 Mb/s		
Type of routing	MS-R	QoS-R	LS-R	MS-R	QoS-R	LS-R	MS-R	QoS-R	LS-R
Delay[ms]	-	-	-	28.71	29.06	10807	31.96	30.25	25895.3
Losses[%]	-	-	-	0	0	80.16	0	0	76.34
Throughput[Mb/s]	-	-	-	3.92	3.42	1.78	8.60	8.60	1.93

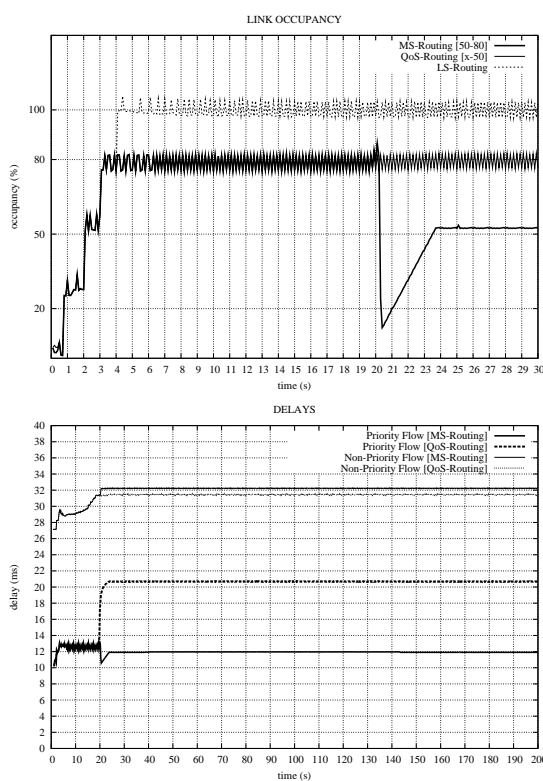


Fig. 4. Longterm behaviour

sharply decreases. If QoS routing is used, when the threshold is crossed every incoming new flow (priority or non-priority) is transmitted through a longer path. Thus, link occupancy is kept near 80% and the delay of priority traffic increases approximately 80%.

5 Conclusions

Existing QoS routing protocols are not able to deal efficiently with service differentiation. The proposed routing protocol provides this kind of support. To perform this, several extensions which provide a solution compatible with traditional routing protocols, with scalability characteristics, have been proposed. Simulation results have shown that priority traffic will achieve better performance and non-priority traffic will suffer less losses. Future work comprises testing the Multi-Service routing in more complex networks; study of other metrics and the integration into an IPv6/MPLS trial platform.

References

- [1] X. Xipeng and M. N. Lionel: Internet QoS: A Big Picture. IEEE Network, March/April (1999).
- [2] S. Chen and K. Nahrstedt: An Overview of Next-Generation High-Speed Networks: Problems and Solutions. IEEE Network, November/December (1998).
- [3] M. Garey and D. Johnson: Computers and Intractability: A Guide to the Theory of NP-completeness. New York: W. H. Freeman ZhuPar95andCo (1979).
- [4] S. Nclakuditi, Z. Zhang and C. H. Du David: On selection of candidate paths for proportional routing. Computer Networks 44 (2004) 79-102.
- [5] A. Khanna and J. Zinky.: The revised ARPANET Routing Metric. Proceedings of SIGCOMM'89, September (1989).
- [6] R. Guérin, S. Kamat, A. Orda, T. Prygienda and D. Williams: QoS Routing Mechanisms and OSPF extensions IETF RFC 2676, August (1999).
- [7] ATM Forum: Private network network interface , Specification Version 1 (PNNI 1.0). March (1996).
- [8] R. Guérin and A. Orda: QoS Based Routing in Networks with Inaccurate Information: Theory and Algorithms Proceedings of IEEE Infocom'97, Japan (1997).
- [9] S. Vutukury and Garvia-Luna-Acheves: A Simple Approximation to Minimum Delay Bridge. Proceedings of ACM SIGCOMM'99, August/September (1999).
- [10] M. Oliveira, B.Melo, G. Quadros and E. Monteiro: Quality of Service Routing in the Differentiated Services Framework Proceedings of SPIE's International Symposium on Voice, Video and Data Communications (Internet III: Quality of Service and Future Directions), Boston, Massachutetts, USA, November 5-8 (2000).
- [11] J. Wang and K. Nahrsted: Hop-by-hop Routing Algorithms for Premium-Class Traffic in DiffServ Networks. Proceedings of IEEE INFOCOM 2002, New York, June (2002).
- [12] K. Nahrstedt and S. Chen: Coexistence of QoS and Best Effort Flows - Routing and Scheduling Proceedings of the 10th IEEE International Workshop on Digital Communications: Multimedia Communications, Ischia, Italy, September (1998).
- [13] Q. Ma and P. Steenkiste: Support Dynamic Inter-Class Resource Sharing: A Multi-Class QoS Routing Algorithm Proceedings of IEEE INFOCOM'99, New York, March (1999).
- [14] E. Crawley, R. Nair, B. Tajagopalan and H. Sandick: A Framework for QoS based routing. IETF RFC 2386, August 1998.
- [15] L. Cidon, R. Rom and Y. Shavitt: Multi-path Routing Combined with Resource Reservation. Proceedings of IEEE INFOCOM'97, Japan, April (1997).
- [16] A. Varela, T. Vazão and G. Arroz: Multi-Service Routing: a New QoS Routing Approach Supporting Service Differentiation. Proceedings of AICT'05, Lisbon, April (2005).