A Class-Pining Election Scheme for Loop-Free QoS Routing

Marilia Curado, Edmundo Monteiro

Abstract— **The main contribution of this paper is the proposal and evaluation of an election approach to provide loop-free class pinning for QoS routing in IP networks. The proposed strategy aims at achieving stability in QoS routing protocols, a situation that is inherently prone to instability. Although the concept of class pinning is inspired in the concept of route pinning, it poses several additional problems, namely, the pinning of paths in hop-by-hop routing and path selection based only on destination address and class of traffic. In hop-by-hop routing, the pinning decision at individual nodes may cause routing loops, since the nodes will stop to have a coherent view of the state of the network. In view of this problem, the proposed approach addresses class pinning for distributed QoS routing schemes guaranteeing that routing loops will not take place.**

Index Terms—**Class-pinning, QoS Routing, Stability**

I. INTRODUCTION

THE main role of QoS routing in IP networks is to dynamically select paths based on information about the state of the network. Therefore, it enables the avoidance of congested paths, contributing to improve application performance and network resource usage. However, the dynamic selection of paths may cause routing instability and network oscillatory behavior, causing the degradation of application performance. In face of this scenario it is necessary to achieve a compromise between the desired adaptability of the protocol and the unwanted instability [1, 2].

Some basic mechanisms to avoid instability in class-based QoS IP networks have been proposed by the authors in previous works [3, 4]. These mechanisms prove to be effective under normal situations but in some circumstances (heavy load and bursty traffic), there is the need of additional measures in order to enhance routing stability characteristics. This objective is fulfilled by the election approach for class-pinning methodology proposed in this paper.

The Class-Pinning (CP) mechanism addresses the instability problem by controlling the instant when a traffic class shifts to a new path. Normally, when the state of the network changes, due to events such as the start of a new flow or a traffic burst, routing messages are sent to all nodes, and new paths are computed. After the calculation, traffic will shift to the less congested paths, leaving the paths currently used. The next time this process occurs, traffic will eventually go back to the original path, and, thus, instability happens. With the class-pinning mechanism, new paths are also computed upon the arrival of routing messages. However, after the computation of the Shortest Path Trees, the weight of the path for each pair destination-class is compared to the weight of the previously installed path. The new path for a destination will only be installed in the routing table if it is significantly better than the path that is currently used by that class.

The class-pinning mechanism proposed is inspired on the route-pinning mechanism of QoS routed flows discussed in RFC 2386 [5]. There are two main differences between route-pinning and class-pinning, one that concerns the routing model and other that concerns routing granularity. Route-pinning is usually used in the context of source routing of flows while class-pinning is used in the context of hop-by-hop routing of traffic aggregates according to destination and class. The distributed routing model used in hop-by-hop routing poses constraints on the design of the class-pinning mechanism, specifically due to the possibility of the occurrence of loops. When a class-pinning decision is taken independently by a single node, the view that all the nodes in the area have of the network will not be consistent. In particular, if the path computation algorithm is based on the Dijkstra algorithm, as in many QoS routing strategies, loops may happen [6]. In order to avoid this problem a new protocol named Class-Pinning Election (CPE) scheme is proposed. The purpose of the CPE scheme is to provide a class-pinning activation mechanism for QoS-aware class-based nodes while protecting the network against routing loops, as will be described in the remaining of this paper.

The rest of the paper is organized as follows. Section 2 describes the proposed CPE scheme. In Section 3 the scheme is evaluated in the Géant European academic network backbone topology. Finally, Section 4 presents some conclusions and directions for future work.

II. THE CPE SCHEME

The objective of the proposed scheme is to provide class-

Manuscript received December 20, 2004. This work was supported in part by the Portuguese Ministry of Science and High Education (MCES), and by European Union FEDER under program POSI (Project QoS II and IPQoS) and under the PhD grant PRAXIS XXI/ BD/13723/97.

Marilia Curado and Edmundo Monteiro are with the Laboratory of Communications and Telematics, CISUC/DEI, Pólo II, Pinhal de Marrocos 3030-290 Coimbra, Portugal (phone: +3512397900; fax: +351239701266; e-mails: marilia@ dei.uc.pt, edmundo@dei.uc.pt).

pinning capabilities in autonomous systems that use a distributed routing model and destination-class routing granularity. The CPE scheme is applied to an environment composed of one Class-Pinning Control Node (CPCN) and a set of Ordinary Nodes (ON). The CPCN is selected among all the nodes in the area, the Ordinary Nodes, to become the responsible for the dynamics of the class-pinning mechanism. The selection of the CPCN is done by a ring-based selection protocol and constitutes the first phase of the CPE scheme.

During the normal execution of the routing algorithm, an ON detects situations when it is suitable to activate the class-pinning mechanism. These situations arise when there isn't a significant difference between the cost of successively computed paths for a certain destination and class. Therefore, when an ON detects such a situation where different next-hops are being successively chosen for the same destination and traffic class without a significant improvement in path cost, it must send a request to the CPCN. The CPCN will then make a decision about the request through a class-pinning election algorithm that will take place only if a significant number of pinning requests have been received. The election process takes into account the characteristics of the ONs that sent class-pinning requests to the CPCN, including the number of interfaces, the capacity of the links and the localization of the node on the area, namely, if it is a central or a border node. This information must be collected by the CPCN and must be kept updated during the operation of the protocol.

A. *Class-Pinning Basics*

The objective of the class-pinning mechanism is to limit the number of path shifts due to dynamic changes in the state of the network. This adaptability is desirable to allow for the choice of the best paths for each traffic class. However, it can originate oscillatory behavior with negative impact on the performance of the routing strategy. The instability problem is tackled with the class-pinning mechanism by forcing the traffic of each class to remain on the same path as long as any alternative path is not significantly better.

The problem of class pinning is to force traffic of each class for a certain destination to use the same next-hop as long as the alternative next-hop is not sufficiently appealing to be worthwhile the change. The factor used to define the level of path cost improvement that must be achieved in order to allow a path shift is the Degree of Significance (*DS*). Whenever this condition is verified, a new next-hop is installed for the corresponding destination and class. The class-pinning mechanism is activated only after the condition has not been satisfied for a certain number of times, defined by the Stability Forecast (*SF*) factor. This approach is used to guarantee that the class-pinning mechanism is only activated when the paths are stable, avoiding thus excessive pinning requests, that would be quickly tear-down.

Figure 1 shows the algorithm used for the activation of the class-pinning mechanism in an ON, where SPT is the shortest path tree for class *i*. The network is represented by a directed graph *G(V,E)*, composed of a set of vertices *V* and a set of edges *E*. The number of vertices of *G* is given by $n = |V|$ and the number of edges is given by $m = |E|$. The weight of the path for destination *j* and class *i* using *nh(i,j)* is *w(i,j)*. The number of times that the class-pinning condition has been verified is represented by n_{cp}.

```
CLASS-PINNING(G, SPT, i, DS, SF)
for j in G
    if ($w_{new}(i, j) < DS * w_{old}(i, j)$)  then
        $n_{cp} \leftarrow 0$
        if (class_pinning = activated)   then
            stop_ class_pinning (i, j)
        endif
    else
        $n_{cp} \leftarrow n_{cp} + 1$
        if ($n_{cp} = SF$)  then
            request_ class_pinning (i, j)
        endif
    endif
endfor
$nh(i, j) \leftarrow nh_{new}(i, j)$
```

Fig. 1. Algorithm for the activation of the class-pinning mechanism.

The algorithm used for the activation of the class-pinning mechanism results in the emission of a pinning request that If there isn't a CPCN in the area, one must be selected by the process described in the next sub-section.

B. *Class Pinning Control Node Selection Protocol*

The selection of the CPCN is done through a *ring-based selection* protocol based on the selection method of the Designated Node used in OSPF [6]. There is however an important difference, since the Designated Node is elected for an IP network while the CPCN must be chosen for an area. When an ON needs to send a class-pinning request and verifies that there isn't any selected CPCN node, it triggers the selection process of the CPCN. This node is called the Requesting Node (RN). The eligibility of nodes to become CPCN is controlled by the Control Node Parameter (*CNP*). Nodes that are not eligible to become CPCN have the *CNP* set to zero, while the others are assigned a positive integer value, with a configured upper bound denominated crp_{max}. If the requesting node has an eligible *CNP* it becomes the CPCN and advertises it to all the other nodes in the area. If the self-selection of the RN is not possible, it starts a ring-based selection process, by scanning nodes that are one-hop neighbours, two-hop neighbours and so on, until either one node becomes CPCN or all the routers have been scanned and none is eligible. The successive sets of neighbours constitute the rings on which the selection process lays.

In the first series of neighbours where there are eligible nodes, all the eligible nodes become self-selected CPCN. Then, all the self-selected nodes flood their information and then each node becomes responsible for the definition of the CPCN. Upon the reception of this message each node decides which node should be the CPCN by choosing the node with the highest CNP. It there is more than one node with the same CNP, the one with the node with highest node identification number is chosen. As soon as there is a ring containing at least one eligible node, the emission of CPCN selection requests stops. As soon as an eligible node is found, the selection process ends and there is the need, at the most, to decide among self-elected nodes, in the radius where the selection process ended. If there isn't any eligible node in the area, the selection is aborted, and the RR informs all the nodes in the area, that will stop the emission of class-pinning requests for a determined period.

The complexity of the selection process depends on the ring where the first eligible node is found. If the RN is eligible the complexity is $\Theta(1)$, and for the other rings the complexity becomes $\Theta(n)$ where n is the number of nodes. After there is an elected CPCN in the area, it is ready to accept class-pinning requests from the ONs and proceed to the class-pinning election process, as will be described in the following section.

C. Class-Pinning Election Algorithm

In the face of the need of introducing pinning characteristics in a hop-by-hop routing strategy, once a node is in a stable situation and wishes to do pinning, there is the need to guarantee that the class-pinning for the specified class and destination is followed by all the nodes in the network. However, if this decision was made based solely in the state of one node, this could result in performance degradation. Thus, in order to avoid this unilateral decision, an election process takes place where a class-pinning request only becomes effective if several requests have been received by the CPCN. Moreover, since nodes in the network have different levels of importance, namely concerning their link capacity, number of interfaces and localization, each node will have a different weight on the voting process.

The class-pinning election process is triggered when the CPCN has received a certain amount of class-pinning requests for destination d and traffic class i, represented by $R(d,i)$. The definition of the amount of requests that triggers the election process depends on the *level of responsiveness* established for each class, ω_i. A very responsive class-pinning mechanism is achieved when $\omega_i = 0$, meaning that the election process will take place as soon as one request arrives, and a strictly responsive situation is achieved when $\omega_i = 1$, meaning that the class-pinning election only occurs when all the nodes have made such a request.

In order to support the election process, nodes are classified by localization, by average link capacity, $\overline{l(v)}$., and number of interfaces (degree of the node), $\gamma(v)$. Concerning localization, nodes are classified as central nodes or as border nodes, according to Definition 1.

Definition 1: A central node is a node whose average distance, in terms of the number of hops, to all the other nodes in the network is smaller than the network radius, thus satisfying $\overline{d_v} < \rho(G)$, where $\overline{d_v}$ is the average distance of a node v to all the other nodes and $\rho(G)$ is the radius of the network represented by graph G. The radius of the network is approximated by $2\rho(G) \geq \Phi(G)$, where $\Phi(G)$ is the diameter of the network. The diameter of a network is the maximum distance between any two nodes.

The parameter that quantifies the localization of a node is the *centrality factor*, $\psi(v)$. A central node has weight 1, and border nodes are assigned a smaller value that decreases exponentially with the distance to the central area of the network graph, as shown in Equation 1.

$$\psi(v) = \begin{cases} 1, & \overline{d_v} \leq \rho(G) \\ \dfrac{1}{e^{(\overline{d_v}-\rho)/\rho}}, & otherwise \end{cases} \qquad (1)$$

The classification of nodes according average link capacity and node degree is done through a combined process based on the (l,γ) plane, where l represents the range of link capacities in the network and γ represents the degree of nodes. The selection of categories on the (l,γ) plane requires the definition of the maximum degree of G, $\Gamma(G)$, and the maximum link capacity, $L_{max}(G)$. The dimension of the categories is controlled by two parameters, one for the node degree range, $\Delta\gamma$, and the other for the link capacity range, Δl. These parameters are a fraction of the maximum corresponding values, by the definition of the degree threshold, T_γ, and the capacity threshold, T_l. Based on the above defined information, nodes are classified in four categories, as shown in Table 1, using two classification styles, one strict and the other relaxed.

TABLE 1.
CATEGORIES OF NODES BASED ON LINK CAPACITY AND NODE DEGREE.

Category	Link capacity	Node degree
I	High	High
II	High	Low/Medium
III	Low/Medium	High
IV	Low/Medium	Low/Medium

The motivation behind the definition of the two styles of classification, strict and relaxed, is due to the need of adapting to networks where there is a broad variety of nodes. In these situations, using a relaxed classification allows for a better distribution of nodes among the categories considered, avoiding an excessive supremacy of high bandwidth or highly connected nodes. When the nodes in the network are more homogeneous, a strict classification style is adequate, since the distribution of weight on the election process is fairly distributed without prejudice of the majority of nodes. Thus, depending on the overall characteristics of the network the more adequate style may be chosen.

Definition 2: Strict classification divides the (L,γ) plane in four rectangular areas, where the upper bounds of categories *I* to *III* correspond to the horizontal and vertical lines on the maximum values for link capacity and node degree in the network, respectively, $L_{max}(G)$ and $\Gamma(G)$. The lower bounds of these categories, which are also the upper bounds of category *IV*, are defined by the horizontal and vertical lines that cross the axes on $L_{max}(G)-\Delta l$ and on $\Gamma(G)-\Delta\gamma$, defining thus the regions of each category.

The relaxed classification style aims at making the election process fairer when the network is more heterogeneous in terms of link capacity and node degree. Therefore, the definition of relaxed categories takes into consideration the average link capacity of the network, $\overline{L(G)}$, and the average node degree of the network, $\overline{\gamma(G)}$.

Definition 3: Relaxed classification divides the (L, γ) plane in four areas confined by the axes and by the horizontal and vertical lines on the maximum values for link capacity and node degree in the network, respectively, $L_{max}(G)$ and $\Gamma(G)$. In the interior, the above defined area is split by the lines that link the points $L_{max}(G)$- Δl and on $\Gamma(G)$- $\Delta\gamma$ to the average node degree, $\overline{\gamma(G)}$, and average link capacity, $\overline{L(G)}$, respectively. More specifically, a node *v* belongs to a category $i\in\{I, II, III, IV\}$, if the point describing the node characteristics $(\overline{l(v)},\gamma(v))$ on the (L, γ) plane is contained on *Area i* (A_i) defined by the four lines that confine each area.

After the two levels of node classification, the final weight that each node *v* will have, W_v, is computed based on the category of the node as defined above, on its centrality, $\psi(v)$, and on the number of class-pinning requests that have been received from that node on the previous period for the same destination and class, $R_v(d,i)$. In order to recompense central nodes that have links with limited amount of capacity and a small node degree, it is introduced a compensation factor, F_c, on the weight computation. These nodes should have a high weight since being central may receive unpredictable traffic flows from a wide range of sources and towards a broad range of destinations and thus their behavior can have an important impact on network performance. Equation 2 shows the function used to compute the weight of a node.

$$W_v(d,i) = \begin{cases} R_v(d,i)^{1/C_v}.\psi(v),\ C_v \in [1,3] \\ R_v(d,i)^{1/C_v}.\psi(v),\ C_v = 4 \wedge \psi(v) < 1 \\ R_v(d,i)^{1/C_v}.\psi(v).F_c,\ C_v = 4 \wedge \psi(v) = 1 \end{cases} \quad (2)$$

The result of the election is obtained by the comparison of the sum of the weight of the nodes that made requests with the sum of the weight of the nodes that did not make any request. If the first sum is higher than the second, the class-pinning process is activated, and the CPCN sends an activation message to all routers. The metric used to evaluate the success of the election process is the Class-Pinning Success Ratio (CPSR), computed by the ratio between the total weight of the routers that made class-pinning requests and the total weight of the other routers.

The evaluation of the mechanisms used to control the class-pinning process and the behavior of the CPE scheme is described in the next section.

III. EVALUATION OF THE CPE SCHEME

This section includes a preliminary analysis of the class-pinning election mechanism on the Géant topology (www.geant.net). The objectives of the analysis are to perform the comparison of the strict and relaxed classification styles and to validate the efficiency of the CPE scheme under different combinations of the configuration parameters, namely the link capacity and node degree thresholds, and the compensation factor. The link capacity and node degree thresholds are varied from 0 to 1 with increases of 0.1 and the compensation factor is varied from 1 to 10. To evaluate the voting process, the requests from 10 different nodes are considered, using two sets of nodes, one with low degree and low average link capacity, the "small" nodes, and the other with high degree and high average link capacity, called "large" nodes. Small nodes have average link capacity of 155Mbps and node degree of 1 or 2, while large nodes have link capacity between 1300 and 8000 Mbps and node degree between 4 and 10, as included in the Géant topology.

Figure 1 shows the number of nodes in each category, when is used strict classification. With small thresholds, most nodes belong to category 4 since there is a prevalence of low degree and low average link capacity nodes. However, when the thresholds increase, the differentiation among nodes becomes visible and some nodes are classified on the other categories. For larger thresholds most nodes are classified in category 1 and will thus have higher weight on the voting process. Figure 1 shows that thresholds in the range [0.4, 0.7] allow for a fair voting process due to a balanced distribution of nodes among categories. In the case of the relaxed classification style, a stronger differentiation of nodes is achieved, since a higher number of nodes is classified in the category with highest weight, as depicted in Figure 2. This ensures a level of differentiation of nodes that have similar characteristics and would be otherwise classified in the same category.

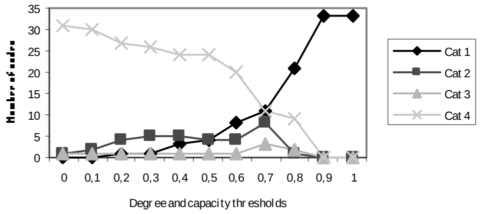

Fig. 1. Result of the classification of nodes with strict style.

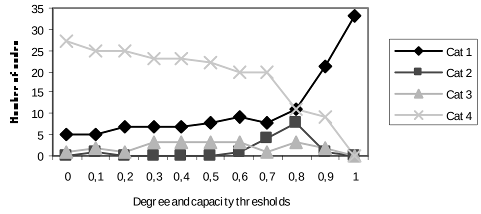

Fig. 2. Classification of nodes with relaxed style.

The result of the voting process for the whole range of degree and capacity thresholds, without the activation of the compensation factor is shown in Figure 3. The most important nodes, Strict Large Node (SLN) and Relaxed Large Node (RLN) have consistently higher weight and are able to control de pinning decision, except when the thresholds are equal to 1, where all nodes belong to the same category. These results show that the voting process has the adequate outcome and favors the most representative nodes, namely, nodes with higher degree and average link capacity. The relaxed classification mode is less sensitive to the dimension of the thresholds, as expected, since the nodes are more balanced among categories.

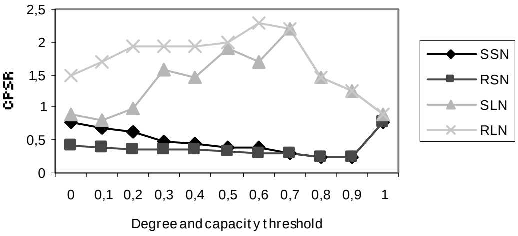

Fig. 3. Election results without compensation factor.

Figure 4 shows the result of the voting process when the compensation factor is 50%. In this case, the less important nodes, Strict Small Node (SSN) and Relaxed Small Node (RSN) outperform slightly the larger nodes and are able to control the pinning decision for almost all thresholds. The activation of the compensation factor is a decision that must be taken at a traffic engineering level, since its role should be limited to particular situations as described above.

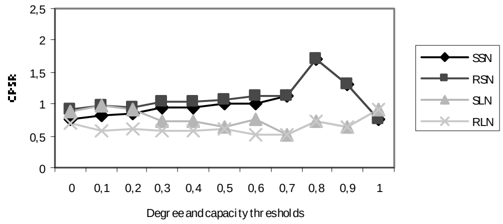

Fig. 4. Election results with compensation factor = 0.5.

IV. CONCLUSION

Stability is a major concern in dynamic routing protocols, and even more important when QoS comes into play. Route pinning is one approach used to limit the instability of routing protocols when the routing decision is done at the flow level and the routing paradigm is source routing. Inspired in this mechanism, a scheme for the pinning of routes in class-based hop-by-hop routing called Class Pinning Election was proposed in this paper.

The CPE scheme has two main components, namely, the detection of the situations when class pinning should be activated and the voting process that decides if the class pinning for a certain class and destination should be enforced. The decision of the voting process is determined by the nodes that made the class-pinning request and the weight of all the nodes in the network. The weight of each node is computed based on the average link capacity of its interfaces, the number of interfaces and the location of the node in the network, namely if it is a central or a border node. The most important nodes, that is, central nodes with more links and higher link capacity, will have higher weight. The occurrence of loops that could arise due to the activation of the class pinning mechanism in individual nodes was addressed by the inclusion of a Class-Pinning Control Node (CPCN) that controls centrally the activation of the class pinning mechanism and guarantees that all the nodes will have a consistent view of the state of the network.

The results have shown that the CPE scheme is able to adequately differentiate among the types of nodes according to the parameters considered and that the election process has the potential to control the activation of class pinning.

ACKNOWLEDGMENT

This work was partially supported by the Portuguese Ministry of Science and High Education (MCES), and by European Union FEDER under program POSI (Project QoS II and IPQoS) and under the PhD grant PRAXIS XXI/ BD/13723/97.

REFERENCES

[1] A. Shaikh, A. Varma, L. Kalampoukas and R. Dube, “Routing stability in congested networks” Proceedings of ACM SIGCOMM’00, August 28-Sepetmber 1, Grand Hotel, Stockholm, Sweden, 2000.
[2] A. Basu, J. Riecke, “Stability Issues in OSPF Routing”, Proceedings of ACM SIGCOMM’01, San Diego, California, USA, August 27-31, 2001.
[3] M. Curado, O. Reis, J. Brito, G. Quadros, E. Monteiro, “Stability and scalability issues in Hop-By-Hop Class-based Routing”, Proceedings of the 2nd International Workshop on QoS in Multiservice IP Networks (QoS-IP 2003), Milano, Italy, February 24-26, 2003.
[4] M. Curado, O. Reis, J. Brito, G. Quadros, E. Monteiro, “Jungle Issues in Class-based Routing”, International Workshop on Multimedia Interactive Protocols and Systems (MIPS 2003), Napoli, November 18-21, 2003.
[5] E. Crawley, R. Nair, B. Rajagopalan, H. Sandick, “A Framework for QoS-based Routing in the Internet”, IETF,RFC 2386, August 1998.
[6] J. Moy, “OSPF Version 2”, IETF,RFC 2328, April 1998.