               Label Switched Path (LSP) Preemption Policies for
                         MPLS Traffic Engineering

Status of This Memo

IESG Note

Abstract

   When the establishment of a higher priority (Traffic Engineering
   Label Switched Path) TE LSP requires the preemption of a set of lower
   priority TE LSPs, a node has to make a local decision to select which
   TE LSPs will be preempted. The preempted LSPs are then rerouted by
   their respective Head-end Label Switch Router (LSR). This document
   presents a flexible policy that can be used to achieve different
   objectives: preempt the lowest priority LSPs; preempt the minimum
   number of LSPs; preempt the set of TE LSPs that provide the closest
   amount of bandwidth to the required bandwidth for the preempting TE
   LSPs (to minimize bandwidth wastage); preempt the LSPs that will have
   the maximum chance to get rerouted. Simulation results are given and

a comparison among several different policies, with respect to
preemption cascading, number of preempted LSPs, priority, wasted
bandwidth and blocking probability is also included.

Table of Contents

1. Motivation

    The IETF Traffic Engineering Working Group has defined the
    requirements and protocol extensions for DiffServ-aware MPLS Traffic
    Engineering (DS-TE) [RFC3564] [RFC4124]. Several Bandwidth
    Constraint models for use with DS-TE have been proposed [RFC4127]
    [RFC4128] [RFC4126] and their performance was analyzed with respect
    to the use of preemption.

    Preemption can be used as a tool to help ensure that high priority
    LSPs can always be routed through relatively favorable paths.
    Preemption can also be used to implement various prioritized access
    policies as well as restoration policies following fault events
    [RFC2702].

    Although not a mandatory attribute in the traditional IP world,
    preemption becomes important in networks using online, distributed
    Constrained Shortest Path First (CSPF) strategies for their Traffic
    Engineering Label Switched Path (TE LSP) path computation to limit
    the impact of bandwidth fragmentation. Moreover, preemption is an
    attractive strategy in an MPLS network in which traffic is treated in
    a differentiated manner and high-importance traffic may be given
    special treatment over lower-importance traffic [DEC-PREP, ATM-PREP].
    Nevertheless, in the DS-TE approach, whose issues and requirements
    are discussed in [RFC3564], the preemption policy is considered an
    important piece on the bandwidth reservation and management puzzle,
    but no preemption strategy is defined. Note that preemption also
    plays an important role in regular MPLS Traffic Engineering
    environments (with a single pool of bandwidth).

    This document proposes a flexible preemption policy that can be
    adjusted in order to give different weight to various preemption
    criteria: priority of LSPs to be preempted, number of LSPs to be
    preempted, amount of bandwidth preempted, blocking probability. The
    implications (cascading effect, bandwidth wastage, priority of
    preempted LSPs) of selecting a certain order of importance for the
    criteria are discussed for the examples given.

2. Introduction

    In [RFC2702], issues and requirements for Traffic Engineering in an
    MPLS network are highlighted. In order to address both traffic-
    oriented and resource-oriented performance objectives, the authors
    point out the need for priority and preemption parameters as Traffic
    Engineering attributes of traffic trunks. The notion of preemption
    and preemption priority is defined in [RFC3272], and preemption
    attributes are defined in [RFC2702] and [RFC3209].

A traffic trunk is defined as an aggregate of traffic flows belonging
to the same class that are placed inside an LSP [RFC3564]. In this
context, preemption is the act of selecting an LSP that will be
removed from a given path in order to give room to another LSP with a
higher priority (lower preemption number). More specifically, the
preemption attributes determine whether an LSP with a certain setup
preemption priority can preempt another LSP with a lower holding
preemption priority from a given path, when there is competition for
available resources. Note that competing for resources is one
situation in which preemption can be triggered, but other situations
may exist, themselves controlled by a policy.

For readability, a number of definitions from [RFC3564] are repeated
here:

Class-Type (CT): The set of Traffic Trunks crossing a link that is
governed by a specific set of Bandwidth constraints. CT is used for
the purposes of link bandwidth allocation, constraint-based routing,
and admission control. A given Traffic Trunk belongs to the same CT
on all links.

TE-Class: A pair of:

i. A Class-Type.

ii. A preemption priority allowed for that Class-Type. This means
that an LSP transporting a Traffic Trunk from that Class-Type can use
that preemption priority as the set-up priority, as the holding
priority, or both.

By definition, there may be more than one TE-Class using the same CT,
as long as each TE-Class uses a different preemption priority. Also,
there may be more than one TE-Class with the same preemption
priority, provided that each TE-Class uses a different CT. The
network administrator may define the TE-Classes in order to support
preemption across CTs, to avoid preemption within a certain CT, or to
avoid preemption completely, when so desired. To ensure coherent
operation, the same TE-Classes must be configured in every Label
Switched Router (LSR) in the DS-TE domain.

As a consequence of a per-TE-Class treatment, the Interior Gateway
Protocol (IGP) needs to advertise separate Traffic Engineering
information for each TE-Class, which consists of the Unreserved
Bandwidth (UB) information [RFC4124]. The UB information will be
used by the routers, checking against the bandwidth constraint model
parameters, to decide whether preemption is needed. Details on how
to calculate the UB are given in [RFC4124].

3. LSP Setup Procedure and Preemption

   A new LSP setup request has two important parameters: bandwidth and
   preemption priority. The set of LSPs to be preempted can be selected
   by optimizing an objective function that represents these two
   parameters, and the number of LSPs to be preempted. More
   specifically, the objective function could be any, or a combination,
   of the following [DEC-PREP, ATM-PREP]:

   * Preempt the LSPs that have the least priority (preemption
     priority). The Quality of Service (QoS) of high priority traffic
     would be better satisfied, and the cascading effect described below
     can be limited.

   * Preempt the least number of LSPs. The number of LSPs that need to
     be rerouted would be lower.

   * Preempt the least amount of bandwidth that still satisfies the
     request. Resource utilization could be improved. The preemption
     of larger TE LSPs (more than requested) by the newly signaled TE
     LSP implies a larger amount of bandwidth to be rerouted, which is
     likely to increase the probability of blocking (inability to find a
     path for some TE LSPs).

   * Preempt LSPs that minimize the blocking probability (risk that
     preempted TE LSP cannot be rerouted).

   After the preemption selection phase is finished, the selected LSPs
   are signaled as preempted and the new LSP is established (if a new
   path satisfying the constraints can be found). The UB information is
   then updated via flooding of an IGP-TE update and/or simply pruning
   the link where preemption occurred. Figure 1 shows a flowchart that
   summarizes how each LSP setup request is treated in a preemption-
   enabled scenario.

```
      LSP Setup Request
    (TE-Class i, bw=r)
               |
               |
               |
               v                  NO
   UB[TE-Class i] >= r ? -------> Reject LSP
                                    Setup and flood an updated IGP-TE
                 |                  LSA/LSP
                 |YES
                 v                  NO
    Preemption Needed ? -------> Setup LSP/Update UB if a threshold is
                 |                          crossed
                 | YES
                 v
        Preemption    ---->      Setup LSP/Reroute Preempted LSPs
        Algorithm                Update UB
```
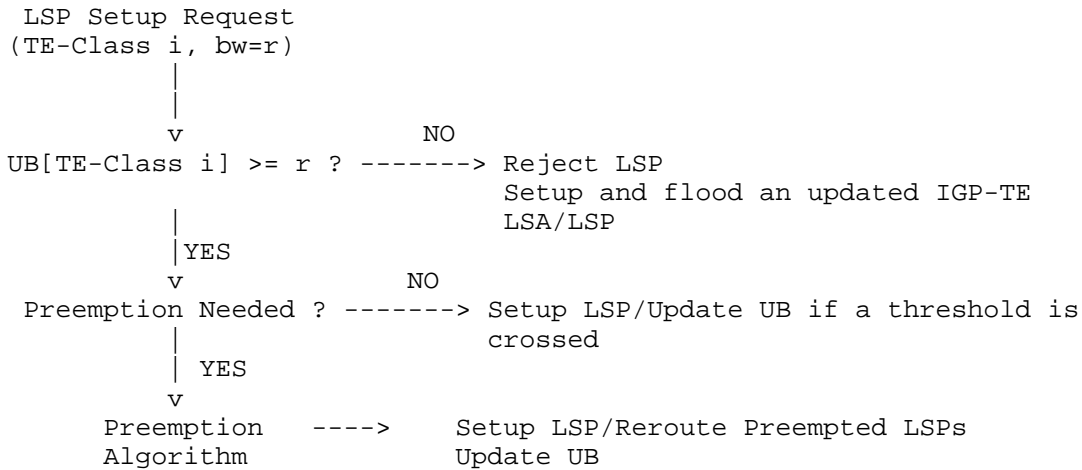
Figure 1: Flowchart for LSP setup procedure.

In [DEC-PREP], the authors propose connection preemption policies
that optimize the discussed criteria in a given order of importance:
number of LSPs, bandwidth, and priority; bandwidth, priority, and
number of LSPs. The novelty in our approach is the use of an
objective function that can be adjusted by the service provider in
order to stress the desired criteria. No particular criteria order
is enforced. Moreover, a new criterion is added to the objective
function: optimize the blocking probability (to minimize the risk
that an LSP is not rerouted after being preempted).

4. Preemption Cascading

   The decision of preempting an LSP may cause other preemptions in the
   network. This is called preemption cascading effect and different
   cascading levels may be achieved by the preemption of a single LSP.
   The cascading levels are defined in the following manner: when an LSP
   is preempted and rerouted without causing any further preemption, the
   cascading is said to be of level zero. However, when a preempted LSP
   is rerouted, and in order to be established in the new route it also
   causes the preemption of other LSPs, the cascading is said to be of
   level 1, and so on.

   Preemption cascading is not desirable and therefore policies that
   minimize it are of interest. Typically, this can result in severe
   network instabilities. In Section 5, a new versatile preemption
   heuristic will be presented. In Section 6, preemption simulation
   results will be discussed and the cascading effect will be analyzed.

5. Preemption Heuristic

5.1. Preempting Resources on a Path

   It is important to note that once a request for an LSP setup arrives,
   each LSR along the TE LSP path checks the available bandwidth on its
   outgoing link. For the links in which the available bandwidth is not
   enough, the preemption policy needs to be activated in order to
   guarantee the end-to-end bandwidth reservation for the new LSP. This
   is a distributed approach, in which every node on the path is
   responsible for running the preemption algorithm and determining
   which LSPs would be preempted in order to fit the new request. A
   distributed approach may not lead to an optimal solution.

   Alternatively, in a centralized approach, a manager entity runs the
   preemption policy and determines the best LSPs to be preempted in
   order to free the required bandwidth in all the links that compose
   the path. The preemption policy would try to select LSPs that
   overlap with the path being considered (preempt a single LSP that
   overlaps with the route versus preempt a single LSP on every link
   that belongs to the route).

   Both centralized and distributed approaches have advantages and
   drawbacks. A centralized approach would be more precise, but require
   that the whole network state be stored and updated accordingly, which
   raises scalability issues. In a network where LSPs are mostly
   static, an offline decision can be made to reroute LSPs and the
   centralized approach could be appropriate. However, in a dynamic
   network in which LSPs are set up and torn down in a frequent manner
   because of new TE LSPs, bandwidth increase, reroute due to failure,
   etc., the correctness of the stored network state could be
   questionable. Moreover, the setup time is generally increased when
   compared to a distributed solution. In this scenario, the
   distributed approach would bring more benefits, even when resulting
   in a non-optimal solution (The gain in optimality of a centralized
   approach compared to a distributed approach depends on many factors:
   network topology, traffic matrix, TE strategy, etc.). A distributed
   approach is also easier to be implemented due to the distributed
   nature of the current Internet protocols.

   Since the current Internet routing protocols are essentially
   distributed, a decentralized approach was selected for the LSP
   preemption policy. The parameters required by the new preemption
   policies are currently available for OSPF and Intermediate System to
   Intermediate System (IS-IS).

5.2. Preemption Heuristic Algorithm

   Consider a request for a new LSP setup with bandwidth b and setup
   preemption priority p. When preemption is needed, due to lack of
   available resources, the preemptable LSPs will be chosen among the
   ones with lower holding preemption priority (higher numerical value)
   in order to fit r=b-Abw(l). The variable r represents the actual
   bandwidth that needs to be preempted (the requested, b, minus the
   available bandwidth on link l: Abw(l)).

   L is the set of active LSPs having a holding preemption priority
   lower (numerically higher) than p. So L is the set of candidates for
   preemption. b(l) is the bandwidth reserved by LSP l in L, expressed
   in bandwidth units, and p(l) is the holding preemption priority of
   LSP l.

   In order to represent a cost for each preemption priority, an
   associated cost y(l) inversely related to the holding preemption
   priority p(l) is defined. For simplicity, a linear relation
   y(l)=8-p(l) is chosen. y is a cost vector with L components, y(l). b
   is a reserved bandwidth vector with dimension L, and components b(l).

   Concerning the objective function, four main objectives can be
   reached in the selection of preempted LSPs:

   * minimize the priority of preempted LSPs,
   * minimize the number of preempted LSPs,
   * minimize the preempted bandwidth,
   * minimize the blocking probability.

   To have the widest choice on the overall objective that each service
   provider needs to achieve, the following equation was defined (for
   simplicity chosen as a weighted sum of the above mentioned criteria):

   $H(l)= \alpha \, y(l) + \beta \, 1/b(l) + \gamma \, (b(l)-r)^2 + \theta \, b(l)$

   In this equation:

   - alpha y(l) captures the cost of preempting high priority LSPs.

   - beta 1/b(l) penalizes the preemption of low bandwidth LSPs,
     capturing the cost of preempting a large number of LSPs.

   - gamma (b(l)-r)^2 captures the cost of preemption of LSPs that are
     much larger or much smaller than r.

   - theta b(l) captures the cost of preempting large LSPs.

Coefficients alpha, beta, gamma, and theta can be chosen to emphasize one or more components of H.

The coefficient theta is defined such that theta = 0 if gamma > 0. This is because when trying to minimize the blocking probability of preempted LSPs, the heuristic gives preference to preempting several small LSPs (therefore gamma, which is the weight for minimizing the preempted bandwidth enforcing the selection of LSPs with similar amount of bandwidth as the requested, needs to be set as zero). The selection of several small LSPs in a normally loaded portion of the network will increase the chance that such LSPs are successfully rerouted. Moreover, the selection of several small LSPs may not imply preempting much more than the required bandwidth (resulting in low-bandwidth wastage), as it will be seen in the discussed examples. When preemption is to happen in a heavy loaded portion of the network, to minimize blocking probability, the heuristic will select fewer LSPs for preemption in order to increase the chance of rerouting.

H is calculated for each LSP in L. The LSPs to be preempted are chosen as the ones with smaller H that add enough bandwidth to accommodate r. When sorting LSPs by H, LSPs with the same value for H are ordered by bandwidth b, in increasing order. For each LSP with repeated H, the algorithm checks whether the bandwidth b assigned to only that LSP is enough to satisfy r. If there is no such LSP, it checks whether the bandwidth of each of those LSPs added to the previously preempted LSPs' bandwidth is enough to satisfy r. If that is not true for any LSP in that repeated H-value sequence, the algorithm preempts the LSP that has the larger amount of bandwidth in the sequence, and keeps preempting in decreasing order of b until r is satisfied or the sequence is finished. If the sequence is finished and r is not satisfied, the algorithm again selects LSPs to be preempted based on an increasing order of H. More details on the algorithm are given in [PREEMPTION].

When the objective is to minimize blocking, the heuristic will follow two options on how to calculate H:

* If the link in which preemption is to happen is normally loaded, several small LSPs will be selected for preemption using H(l)= alpha y(l) + theta b(l).

* If the link is overloaded, few LSPs are selected using H(l)= alpha y(l) + beta 1/b(l).

6. Examples

6.1. Simple Case: Single Link

   We first consider a very simple case, in which the path considered
   for preemption is composed by a single hop. The objective of this
   example is to illustrate how the heuristic works. In the next
   section, we will study a more complex case in which the preemption
   policies are being tested on a network.

   Consider a link with 16 LSPs with reserved bandwidth b in Mbps,
   preemption holding priority p, and cost y, as shown in Table 1. In
   this example, 8 TE-Classes are active. The preemption here is being
   performed on a single link as an illustrative example.

   ```
   --------------------------------------------------------------------
   LSP                     L1    L2    L3    L4    L5    L6    L7    L8
   --------------------------------------------------------------------
   Bandwidth (b)           20    10    60    25    20     1    75    45
   Priority  (p)            1     2     3     4     5     6     7     5
   Cost      (y)            7     6     5     4     3     2     1     3
   --------------------------------------------------------------------
   LSP                     L9   L10   L11   L12   L13   L14   L15   L16
   --------------------------------------------------------------------
   Bandwidth (b)          100     5    40    85    50    20    70    25
   Priority  (p)            3     6     4     5     2     3     4     7
   Cost      (y)            5     2     4     3     6     5     4     1
   --------------------------------------------------------------------
   ```
   Table 1: LSPs in the considered link.

   A request for an LSP establishment arrives with r=175 Mbps and p=0
   (highest possible priority, which implies that all LSPs with p>0 in
   Table 1 will be considered when running the algorithm). Assume
   Abw(l)=0.

   If priority is the only important criterion, the network operator
   configures alpha=1, beta=gamma=theta=0. In this case, LSPs L6, L7,
   L10, L12, and L16 are selected for preemption, freeing 191 bandwidth
   units to establish the high-priority LSP. Note that 5 LSPs were
   preempted, but all with a priority level between 5 and 7.

   In a network in which rerouting is an expensive task to perform (and
   the number of rerouted TE LSPs should be as small as possible), one
   might prefer to set beta=1 and alpha=gamma=theta=0. LSPs L9 and L12
   would then be selected for preemption, adding up to 185 bandwidth
   units (less wastage than the previous case). The priorities of the
   selected LSPs are 3 and 5 (which means that they might themselves
   preempt some other LSPs when rerouted).

Suppose the network operator decides that it is more appropriate to configure alpha=1, beta=10, gamma=0, theta=0 (the parameters were set to values that would balance the weight of each component, namely priority and number, in the cost function), because in this network rerouting is very expensive, LSP priority is important, but bandwidth is not a critical issue. In this case, LSPs L7, L12, and L16 are selected for preemption. This configuration results in a smaller number of preempted LSPs when compared to the first case, and the priority levels are kept between 5 and 7.

To take into account the number of LSPs preempted, the preemption priority, and the amount of bandwidth preempted, the network operator may set alpha > 0, beta > 0, and gamma > 0. To achieve a balance among the three components, the parameters need to be normalized. Aiming for a balance, the parameters could be set as alpha=1, beta=10 (bringing the term $1/b(l)$ closer to the other parameters), and gamma=0.001 (bringing the value of the term $(b(l)-r)^2$ closer to the other parameters). LSPs L7 and L9 are selected for preemption, resulting in exactly 175 bandwidth units and with priorities 3 and 7 (note that less LSP are preempted but they have a higher priority which may result in a cascading effect).

If the minimization of the blocking probability is the criterion of most interest, the cost function could be configured with theta=1, alpha=beta=gamma=0. In that case, several small LSPs are selected for preemption: LSPs L2, L4, L5, L6, L7, L10, L14, and L16. Their preemption will free 181 Mbps in this link, and because the selected LSPs have small bandwidth requirement there is a good chance that each of them will find a new route in the network.

From the above example, it can be observed that when the priority was the highest concern and the number of preempted LSPs was not an issue, 5 LSPs with the lowest priority were selected for preemption. When only the number of LSPs was an issue, the minimum number of LSPs was selected for preemption: 2, but the priority was higher than in the previous case. When priority and number were important factors and a possible waste of bandwidth was not an issue, 3 LSPs were selected, adding more bandwidth than requested, but still with low preemption priority. When considering all the parameters but the blocking probability, the smallest set of LSP was selected, 2, adding just enough bandwidth, 175 Mbps, and with priority levels 3 and 7.

When the blocking probability was the criterion of interest, several (8) small LSPs were preempted. The bandwidth wastage is low, but the number of rerouting events will increase. Given the bandwidth requirement of the preempted LSPs, it is expected that the chances of finding a new route for each LSP will be high.

6.2. Network Case

   For these experiments, we consider a 150 nodes topology with an
   average network connectivity of 3. 10% of the nodes in the topology
   have a degree of connectivity of 6. 10% of the links are OC3, 70% are
   OC48, and 20% are OC192.

   Two classes of TE LSPs are in use: Voice LSPs and Data Internet/VPN
   LSPs. For each class of TE LSP, the set of preemptions (and the
   proportion of LSPs for each preemption) and the size distributions
   are as follows (a total of T LSPs is considered):

   T: total number of TE LSPs in the network (T = 18,306 LSPs)

   Voice:

   Number: 20% of T
   Preemption: 0, 1 and 2
   Size: uniform distribution between 30M and 50M

   Internet/VPN TE:

   Number: 4% of T
   Preemption: 3
   Size: uniform distribution between 20M and 50M

   Number: 8% of T
   Preemption 4
   Size: uniform distribution between 15M and 40M

   Number: 8% of T
   Preemption 5
   Size: uniform distribution between 10M and 20M

   Number: 20% of T
   Preemption 6
   Size: uniform distribution between 1M and 20M

   Number: 40% of T
   Preemption 7
   Size: uniform distribution between 1K and 1M

   LSPs are set up mainly due to network failure: a link or a node
   failed and LSPs are rerouted.

The network failure events were simulated with two functions:

- Constant: 1 failure chosen randomly among the set of links every 1
  hour.

- Poisson process with interarrival average = 1 hour.

Table 2 shows the results for simulations with constant failure. The
simulations were run with the preemption heuristic configured to
balance different criteria (left side of the table), and then with
different preemption policies that consider the criteria in a given
order of importance rather than balancing them (right side of the
table).

The proposed heuristic was configured to balance the following
criteria:

HPB: The heuristic with priority and bandwidth wastage as the most
important criteria (alpha=10, beta=0, gamma=0.001, theta=0).

HBlock: The heuristic considering the minimization of blocking
probability (normal load links: alpha=1, beta=0, gamma=0, theta=0.01)
(heavy load links: alpha=1, beta=10).

HNB: The heuristic with number of preemptions and wasted bandwidth in
consideration (alpha=0, beta=10, gamma=0.001, theta=0).

Other algorithms that consider the criteria in a given order of
importance:

P: Sorts candidate LSPs by priority only.

PN: Sorts the LSPs by priority, and for cases in which the priority
is the same, orders those LSPs by decreasing bandwidth (selects
larger LSPs for preemption in order to minimize number of preempted
LSPs).

PB: Sorts the LSPs by priority, and for LSPs with the same priority,
sorts those by increasing bandwidth (select smaller LSPs in order to
reduce bandwidth wastage).

	Heuristic	Other algorithms
	HPB	HBlock	HNB	P	PN	PB
Need to be Rerouted	532	532	532	532	532	532
Preempted	612	483	619	504	477	598
Rerouted Blocked	467|76% 145|24%	341|73% 130|27%	475|77% 144|23%	347|69% 157|31%	335|70% 142|30%	436|73% 162|27%
Max Cascading	4.5	2	5	2.75	2	2.75
Wasted Bandwidth AVR (Mbps) Worst Case(Mbps)	6638 35321	532 26010	6479 36809	8247 28501	8955 31406	6832 23449
Priority Average Worst Case	6 1.5	6.5 3.8	5.8 1.2	6.6 3.8	6.6 3.8	6.6 3.8
Extra Hops Average Worst Case	0.23 3.25	0.25 3	0.22 3.25	0.25 3	0.25 3	0.23 2.75

Table 2: Simulation results for constant network failure:
          1 random failure every hour.

From Table 2, we can conclude that among the heuristic (HPB, HBlock,
HNB) results, HBlock resulted in the smaller number of LSPs being
preempted. More importantly, it also resulted in an overall smaller
rejection rate and smaller average wasted bandwidth (and second
overall smaller worst-case wasted bandwidth.)

Although HBlock does not try to minimize the number of preempted
LSPs, it ends up doing so, because it preempts LSPs with lower
priority mostly, and therefore it does not propagate cascading much
further. Cascading was the overall lowest (preemption caused at most
two levels of preemption, which was also the case for the policy PN).
The average and worst preemption priority was very satisfactory
(preempting mostly lowest-priority LSPs, like the other algorithms P,
PN, and PB).

When HPB was in use, more LSPs were preempted as a consequence of the
higher cascading effect. That is due to the heuristic's choice of
preempting LSPs that are very similar in bandwidth size to the

bandwidth size of the preemptor LSP (which can result in preempting a
higher priority LSP and therefore causing cascading). The wasted
bandwidth was reduced when compared to the other algorithms (P, PN,
PB).

When HNB was used, cascading was higher than the other cases, due to
the fact that LSPs with higher priority could be preempted. When
compared to P, PN, or PB, the heuristic HNB preempted more LSPs (in
fact, it preempted the largest number of LSPs overall, clearly
showing the cascading effect), but the average wasted bandwidth was
smaller, although not as small as HBlock's (the HNB heuristic tries
to preempt a single LSP, meaning it will preempt LSPs that have a
reserved bandwidth similar to the actual bandwidth needed. The
algorithm is not always successful, because such a match may not
exist, and in that case, the wasted bandwidth could be high). The
preempted priority was the highest on average and worse case, which
also shows why the cascading level was also the highest (the
heuristic tries to select LSPs for preemption without looking at
their priority levels). In summary, this policy resulted in a poor
performance.

Policy PN resulted in the small number of preempted LSPs overall and
small number of LSPs not successfully rerouted. Cascading is low,
but bandwidth wastage is very high (overall highest bandwidth
wastage). Moreover, in several cases in which rerouting happened on
portions of the network that were underloaded, the heuristic HBlock
preempted a smaller number of LSPs than PN.

Policy P selects a larger number of LSPs (when compared to PN) with
low priority for preemption, and therefore it is able to successfully
reroute less LSPs when compared to HBlock, HPB, HNB, or PN. The
bandwidth wastage is also higher when compared to any of the
heuristic results or to PB, and it could be worse if the network had
LSPs with a low priority and large bandwidth, which is not the case.

Policy PB, when compared to PN, resulted in a larger number of
preempted LSPs and an overall larger number of blocked LSPs (not
rerouted), due to preemption. Cascading was slightly higher. Since
the selected LSPs have low priority, they are not able to preempt
much further and are blocked when the links are congested. Bandwidth
wastage was smaller since the policy tries to minimize wastage, and
preempted priority was about the same on average and worst case.

The simulation results show that when preemption is based on
priority, cascading is not critical since the preempted LSPs will not
be able to propagate preemption much further. When the number of
LSPs is considered, fewer LSPs are preempted and the chances of
rerouting increases. When bandwidth wastage is considered, smaller

LSPs are preempted in each link and the wasted bandwidth is low. The
heuristic seems to combine these features, yielding the best results,
especially in the case of blocking probability. When the heuristic
was configured to minimize blocking probability (HBlock), small LSPs
with low priority were selected for preemption on normally loaded
links and fewer (larger) LSPs with low priority were selected on
congested links. Due to their low priority, cascading was not an
issue. Several LSPs were selected for preemption, but the rate of
LSPs that were not successfully rerouted was the lowest. Since the
LSPs are small, it is easier to find a new route in the network.
When selecting LSPs on a congested link, fewer larger LSPs are
selected improving load balance. Moreover, the bandwidth wastage was
the overall lowest. In summary, the heuristic is very flexible and
can be configured according to the network provider's best interest
regarding the considered criteria.

For several cases, the failure of a link resulted in no preemption at
all (all LSPs were able to find an alternate path in the network) or
resulted in preemption of very few LSPs and subsequent successfully
rerouting of the same with no cascading effect.

It is also important to note that for all policies in use, the number
of extra hops when LSPs are rerouted was not critical, showing that
preempted LSPs can be rerouted on a path with the same length or a
path that is slightly longer in number of hops.

7. Security Considerations

The practice described in this document does not raise specific
security issues beyond those of existing TE.

8. Acknowledgements

We would like to acknowledge the input and helpful comments from
Francois Le Faucheur (Cisco Systems) and George Uhl (Swales
Aerospace).

9. Informative References

    [ATM-PREP]   Poretsky, S. and Gannon, T., "An Algorithm for
                    Connection Precedence and Preemption in Asynchronous
                    Transfer Mode (ATM) Networks", Proceedings of IEEE ICC
                    1998.

    [DEC-PREP]   Peyravian, M. and Kshemkalyani, A. D. , "Decentralized
                    Network Connection Preemption Algorithms", Computer
                    Networks and ISDN Systems, vol. 30 (11), pp. 1029-1043,
                    June 1998.

    [PREEMPTION]   de Oliveira, J. C. et al., "A New Preemption Policy for
                    DiffServ-Aware Traffic Engineering to Minimize
                    Rerouting", Proceedings of IEEE INFOCOM 2002.

    [RFC2702]   Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and
                    J. McManus, "Requirements for Traffic Engineering Over
                    MPLS", RFC 2702, September 1999.

    [RFC3209]   Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan,
                    V., and G. Swallow, "RSVP-TE: Extensions to RSVP for
                    LSP Tunnels", RFC 3209, December 2001.

    [RFC3272]   Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X.
                    Xiao, "Overview and Principles of Internet Traffic
                    Engineering", RFC 3272, May 2002.

    [RFC3564]   Le Faucheur, F. and W. Lai, "Requirements for Support
                    of Differentiated Services-aware MPLS Traffic
                    Engineering", RFC 3564, July 2003.

    [RFC4124]   Le Faucheur, F., "Protocol Extensions for Support of
                    Diffserv-aware MPLS Traffic Engineering", RFC 4124,
                    June 2005.

    [RFC4126]   Ash, J., "Max Allocation with Reservation Bandwidth
                    Constraints Model for Diffserv-aware MPLS Traffic
                    Engineering & Performance Comparisons", RFC 4126,
                    June 2005.

    [RFC4127]   Le Faucheur, F., "Russian Dolls Bandwidth Constraints
                    Model for Diffserv-aware MPLS Traffic Engineering",
                    RFC 4127, June 2005.

   [RFC4128]   Lai, W., "Bandwidth Constraints Models for
                   Differentiated Services (Diffserv)-aware MPLS Traffic
                   Engineering: Performance Evaluation", RFC 4128,
                   June 2005.

Authors' Addresses

   Jaudelice C. de Oliveira (editor)
   Drexel University
   3141 Chestnut Street (ECE Dept.)
   Philadelphia, PA 19104
   USA

   EMail: jau@ece.drexel.edu

   JP Vasseur (editor)
   Cisco Systems, Inc.
   1414 Massachusetts Avenue
   Boxborough, MA 01719
   USA

   EMail: jpv@cisco.com

   Leonardo Chen
   Verizon Laboratories
   40 Sylvan Rd. LA0MS55
   Waltham, MA 02451
   USA

   EMail: leonardo.c.chen@verizon.com

   Caterina Scoglio
   Kansas State University
   2061 Rathbone Hall
   Manhattan, Kansas 66506-5204
   USA

   EMail: caterina@eece.ksu.edu