

PCEP extensions to support interdomain TED-to-TED feedback

M. Domínguez-Dorado*, J. L. González-Sánchez*, J. Domingo-Pascual*, F. J. Rodríguez-Pérez*, J. Carmona-Murillo*
* University of Extremadura. Avda. de la Universidad s/n. Cáceres, SPAIN.
* Technical University of Catalunya. Jordi Girona, 1-3. Barcelona. SPAIN.

mdomdor@unex.es, jlgs@unex.es, jordi.domingo@ac.upc.edu, fjrodri@unex.es, jcarmur@unex.es

Abstract

The Path Computation Element (PCE) architecture provides enough functionality to compute paths that span multiple domains in Multiprotocol Label Switching (MPLS) and Generalized Multiprotocol Label Switching (GMPLS) networks. This kind of path computations is achieved via the cooperation of PCEs located in adjacent domains. The PCE that start the path computation chooses its foreign peer PCE following a PCE selection mechanism that could take into account the state of the network and its resources. This mechanism is very important and relevant in the overall time taken to compute the complete end-to-end path.

In this work, we contribute with a set of PCE communication Protocol (PCEP) extensions that make possible the exchange of abstract TE information between PCEs located in adjacent domains, avoiding topological disclosures. Having this information in its TED (Traffic Engineering Database) a PCE is able to choose accurately the preferred cooperating PCE minimizing the time needed to compute an interdomain path and taking into account the state of the network.

Keywords

Interdomain routing, MPLS-TE, PCE architecture, PCEP, PCE selection.

1. Introduction

Today, computing routes in the Internet is a hard task; the routes have to fulfil a growing set of constraints that are becoming, at the same time, more and more complex. For example, a route could have to provide a given amount of bandwidth, balance the load of the network or minimize concurrently the residual bandwidth in all the link of the topology. To accomplish these constraints, we can take advantage of TE (Traffic Engineering) [1]. It can be understood as the capability offered by some technologies of monitoring, measuring, managing and modifying the behaviour of operative networks to do their best and to provide the expected QoS (Quality of Service) [2] to the circulating flows. Using other words, having TE enables us to adjust the resources of a physical topology to make the existing services operate finely.

Instead of a single protocol, TE in the Internet is provided via a set of extensions of existing protocols and auxiliary technologies. Thanks to these extensions, ISPs (Internet Service Providers) are able to satisfy the required constraints when computing the paths over their own domains, and this is the reason why it is expected that an ISP supporting TE operation can assure the quality of their services better than an ISP lacking of TE technologies.

1.1. Importance of PCE-based traffic engineering

The PCE (Path Computation Element) architecture [3] allows applying TE in MPLS (Multiprotocol Label Switching) [4] and GMPLS (Generalized Multiprotocol Label Switching) [5] networks, in both intradomain and interdomain

environments [6]. One of the main goals of this architecture is to detach the path computation capability from the nodes that usually do this task. In this fashion, they can be less complex and expensive. At the same time, specialized nodes called PCEs are located in the network to receive the requests of nodes needing path computation services. These PCEs can be improved with sophisticated routing capabilities in order to compute routes bearing in mind TE constraints.

A PCE is able to compute LSPs (Label Switched Paths) over its own domain since it has a complete topological visibility. Nevertheless, when the path to be computed exceeds the local domain, the PCE has to work together with PCEs in other domains to forward the requests to them and join the path segments they compute. In this circumstance, it is achievable for the PCE to have some potential PCEs to cooperate with. The process used to select the preferred one is important in the overall time taken to compute the interdomain route. There are not well-accepted mechanisms to attain a precise PCE selection yet. Therefore, this is still an open topic [7].

1.2. Contribution and plan of this paper

The general aim of our work is to help in the development of the PCE architecture. In [8] we carried out a profoundly discussion about the TE information to be exchanged and now, in this paper, we continue our own work developing a mechanism to make the abstract TE information arrive at other PCEs: a set of PCEP extensions that allow a PCE to distribute its abstract TE information to others.

The rest of this work is organized as follows. In the second section, we do a brief introduction of PCE architecture to focus the scope of our work. In the third section, we summarize the key aspects of our previous work related to the nature and the amount of the TE information to be shared in interdomain environments. The fourth section shows the format, meaning, operation and analysis of a set of PCEP extensions we have designed to transmit TE information between neighbouring domains. To finish, fifth section shows up our conclusions and the work planned for the near future.

2. PCE architecture: a succinct description

In the next paragraphs, we present a short description of the PCE architecture.

2.1. PCE basic architecture

PCE architecture is being developed now so, most of the RFCs (Request for Comments) published by the IETF are definitions and general requirements of the architecture. The most basic PCE architecture must have at least three key elements (Fig. 1). The PCE is the element in charge of paths computation. The PCC (Path Computation Client) is the node that will request the PCE for a path computation; and the PCEP (Path Computation Element communication Protocol) [8], [10], [11], is the communication protocol through which PCEs and PCCs communicate each other. In general, LER (Label Edge Router) will act as PCC since they are access points to the MPLS network.

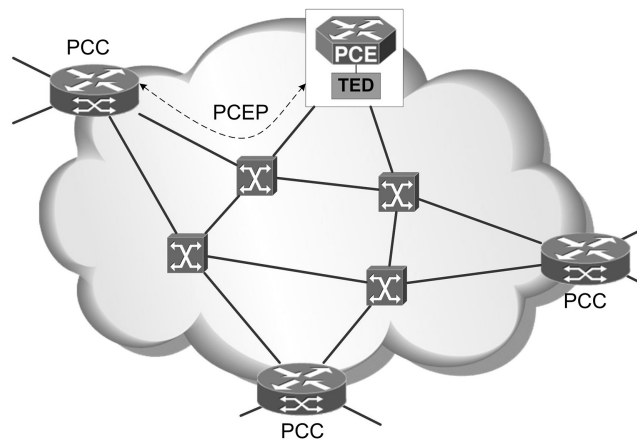


Fig. 1. The most basic PCE architecture

Although at first glance it might appear to be a very simple model, some difficulties come up when integrating PCE in a domain based on legacy technologies. Researchers around the world are managing some of them, e.g. the relationship of PCE, IGP and EGP (Exterior Gateway Protocol) or the way that existing protocols feed PCEs database with traffic engineering information.

In the PCE architecture, a given MPLS domain must have one or more PCE elements responsible of computing LSPs inside it. Each node that wants to start an LSP establishment will have to proceed as a PCC; thus, at least LER nodes must act as PCC because they have the responsibility of establishing LSPs in the interior of the domain. Besides them, other intermediate nodes could need to work as PCC if they are currently concerned in local path restoration mechanisms and need to compute routes.

As one flow arrives to the ingress LER, that LER will act as a PCC and ask the PCE for a path computation from itself to the egress LER. In order to do it, it will use PCEP. The request will incorporate a set of constraints the PCE must take into account when computing the path. It will calculate the route based on the information included in its TED (Traffic Engineering Database), a database that contains the link-state graph and any other information that can be of helpfulness. Each PCE has a TED that is updated regularly by IGPs or by any other method that could be defined (at this moment this is a hot topic [12]). Once the PCE has computed the requested path, it will use PCEP once more to send back a response to the ingress LER/PCC that sent the initial request.

In the requirements description of PCEP, it is specified that a computed path, included in the reply to a LER/PCC, must be directly mapped into an RSVP-TE (Resource Reservation Protocol – Traffic Engineering) ERO [13] (Explicit Routing Object) object so, the LER/PCC is able to start the LSP establishment using RSVP-TE and that ERO object.

This is the most basic operation mode of the PCE architecture (simple path computation), but the PCE architecture allows the existence of other situations, more complex to coordinate and set up, as we can see in Fig. 2. e.g., more than one PCE element can be allocated within a domain, each one of them in charge of computing complete LSPs. In that case, a given LER/PCC will have the chance of choosing the PCE that fits better to its necessities. That situation obliges the LER/PCC to be conscious of the capabilities of each PCE in order to accomplish a selection process based on a reasonable criterion. Another example is the one where there are more than a single PCE inside a domain and each one of them has only the capability of computing path segments related to a specific area of the network (multiple path computation). In that case, these PCEs will have the compulsion of cooperating to compute each own segment, assemble them and give the initiating LER/PCC the path it asked for. For that cause, there are some situations where a PCE has to act as a PCC in face of other PCEs.

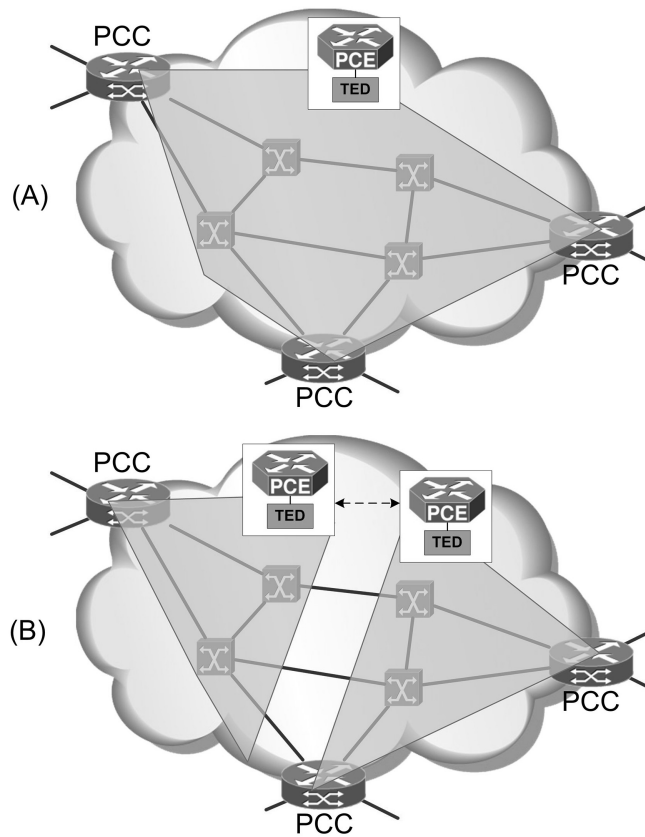


Fig. 2. Single (A) and multiple (B) path computation

2.2. Interdomain PCE architecture.

The most problematic situation happens when the path requested by the LER/PCC exceeds the limit of the local domain. In this case, the PCE architecture should have mechanisms to overcome the traditional interdomain routing problems in relation to traffic engineering and MPLS [14], [15]: partial information about the topology, lack of information about traffic engineering, policy-based routing, uniqueness of routes, security, network recovery, resilience or disclosures of the domain topology.

There is a well-structured operation mode in the interdomain PCE architecture where interdomain cooperation is needed [16], [17]; and it is very similar to the operation mode in interior environments. First, a PCE belonging to a given domain has to know about the existence of other PCEs in its neighbourhood. There are some proposals that supply mechanisms to discover, dynamically and automatically, PCE elements in neighbours domains (most of them has to be considered still as work in progress); for example, those defined in [18], [19], [20], [21] or [22], all of them trying to fulfil the requirements expressed in [23] and most of them following similar principles (Fig. 3).

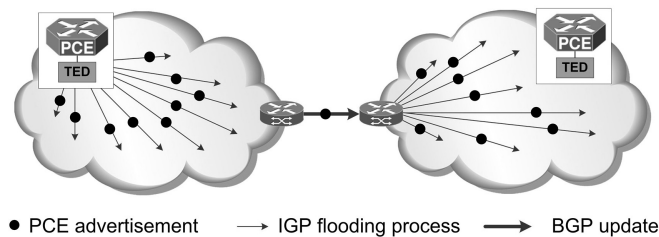


Fig. 3. Generic Interdomain PCE discovery mechanism

Once the PCE has discovered the existing PCE surrounding its own domain, it will have a database containing a set of potential peer PCEs and, therefore, at the moment of receiving an interdomain LSP computation request it must perform the interdomain PCE selection process. Thus, the PCE chooses the foreign PCE that could help it, as a partner, in the overall LSP computation process. To do this, the PCE will retrieve information about the capabilities of all existing PCEs. This information is usually provided together with the discovery advertisement by the interdomain discovery methods.

When the selection process finishes, the initiating PCE contacts the selected PCE via PCEP as usual. The rest of the PCE operation is similar to the traditional PCE operation except for the fact that the PCE must trust the foreign PCE to compute the part of LSP that cross this foreign domain fulfilling the same required set of constraints. The previously mentioned process will be repeated downstream, domain by domain, until the target domain is reached.

2.3. PCE selection. An open issue in PCE architecture

There are still some key aspects to solve before having a valid interdomain PCE architecture. Although several efforts are being made to develop an interdomain discovery mechanism, there is not a well-accepted technique yet; and things are even worse in relation to the interdomain PCE selection process where much more work is needed.

Let us see an explanatory example to understand the difficulties and the impact of selecting the correct peer PCE in interdomain environments. In Fig. 4, we can see an interdomain system where a PCE of the local domain has to compute an interdomain LSP to the target domain. As it has four adjacent domains (D1-D4), it will discover potential peer PCEs in all these domains. There are significant dissimilarities of choosing the PCE of D1, D2, D3 or D4 as a peer and it could have a great impact in the overall time taken to compute the LSP completely. In the example, the best path (and the only one) crosses links A, B, C, and it is computed via D1. Any other selection will result in a negative path computation; anyway, at the end, the correct path will be computed, although several failed attempts will be necessary and the delay in the path computation will be greater.

When two PCE cooperates in a path computation, it is needed an amount of time we call T_{PCEP} that includes the time taken by the computation process of each PCE and the time needed to communicate both PCE via PCEP protocol (generally greater). In general, the process of computing an interdomain LSP in the PCE architecture is proportional to the number of domains the computed LSP has to traverse and takes $n \cdot T_{PCEP}$. Consequently, the way to decrease the overall time taken to compute the complete LSP is to minimize n (the number of interdomain PCE collaborations needed, which depends on the number of involved domains). However, this is not an easy task, mainly taking into account the partial point of view of the PCE in the local domain.

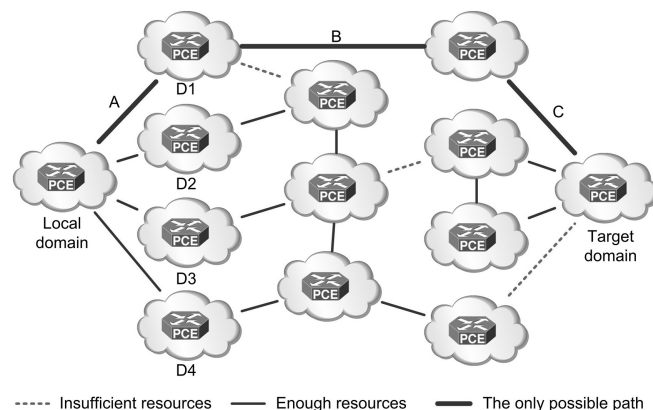


Fig. 4. Explanatory example

Current proposals for interdomain PCE selection [7] includes Round Robin schemes or saving the history of previous requests sent to the foreign PCEs (response time, congestion level...) to avoid those that seem to be busier. Unfortunately, these techniques are not based on information related to the state of the network. Only the state (or the inferred state) of potential peer PCEs is taken into account.

3. Previous work in interdomain PCE selection

In a previous work [8], we proposed a mechanism to provide the PCEs with TE information that is abstracted to ensure the privacy of the involved domains. This information let a given PCE to choose the collaborating foreign PCE accurately, minimizing the number of PCE cooperations needed to compute the complete end-to-end path. Key aspects of this proposal are:

- The TE information used: we studied the TE information supplied by OSPF-TE and ISIS-TE and defined in [24], [25], [26], [27], [29], [30], [31], [32], [33], [34]. After this study, we selected a set of ten TE parameters that were common and comparable for both IGP (Table 1).
- The aggregation functions: in order to avoid privacy disclosures we decided that this TE information was shared in an aggregated form. So it is converted to a coarse-grain TE information, still useful but preserving the privacy of the involved domains. We defined an aggregation function for each selected TE parameter.
- The RI-CUBE (Routing Information CUBE): we designed this data structure to store the aggregated TE information in each PCE and to provide it with advanced query capabilities. Using the RI-CUBE, a PCE is able to predict whether choosing a given PCE as a collaborator will result in a successful path computation or not. It allows the PCE to avoid selecting a peer PCE that seems not to be able of compute the desired LSP.
- The proposal is incremental: this means that not all the PCE has the obligation of supporting this proposal, but the more PCE implements it, the better results will be obtained.

Table 1. Selected set of common and comparable TE parameters.

Description of TE parameter	No. of TE parameter	Format
Maximum bandwidth	1	32 bits IEEE 754 floating point format.
Maximum reservable bandwidth	2	32 bits IEEE 754 floating point format.
Unreserved bandwidth	3	8x32 bits IEEE 754 floating point format. One for each one of the eight existing priority levels.
Maximum available bandwidth for the LSP	4	8x32 bits IEEE 754 floating point format. One for each one of the eight existing priority levels.
Local link protection type	5	1 octet. Values 0x01, 0x02, 0x04, 0x08, 0x10 and 0x20, depending on the selected local link protection
Minimum bandwidth for the LSP	6	32 bits IEEE 754 floating point format.
Maximum transfer unit	7	A 2 octets number $[0-2^{16}]$.
Supports MPLS-TE	8	1 bit. Boolean meaning.
Supports GMPLS	9	1 bit. Boolean meaning.
Switching capability	10	1 octet. Values 1, 2, 3, 4, 51, 100, 151 and 200, depending on the selected switching capability.

Using these mechanisms and techniques, we are able to make a given PCE know the state of the interdomain system, in terms of resources availability, and this fact makes possible the accurate interdomain PCE selection.

4. Proposal of PCEP extensions to exchange TE parameters

Although we know the TE information we should use and the mechanisms we have to apply to aggregate it, we still need a mechanism to make this information reach PCEs in others domains. Our aim is to provide the PCE with enough TE information or visibility, related to the resources of the overall interdomain system, to perform an interdomain PCE selection based on it. Each PCE is in charge of abstracting the TE information related to its own domain. This abstraction process makes the domain appear as a node where crossing it from an entry to an exit point has a cost and an amount of available resources.

Since topology disclosures have to be avoided, this abstract information should not be shared before aggregating it with the information that the surrounding foreign PCEs send to the local one and vice versa. To achieve the TE information exchange, we need a communication protocol that allow the information to be shared between adjacent domains (Fig. 5) and, this way, all the information is grouped and can be spread to others, safely. Thus, the shared information is not related to a single domain but to the complete path to reach the target domain (taking in mind a hypothetical LSP computation request).

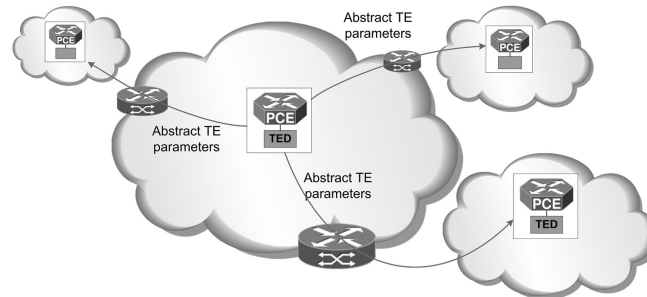


Fig. 5. Interdomain TE parameters exchange

4.1. I3TFP: Interdomain TED-to-TED feedback protocol

In this work, we design a communication protocol called I3TFP (Interdomain TED-to-TED Feedback Protocol) that will be in charge of exchanging aggregated TE information between adjacent domains. It takes advantage of the extension capabilities offered by PCEP and it is, in fact, a set of PCEP extensions; so, we can understand it as a PCEP sub protocol. A TED can be fed by IGP or by any other mechanism that could be designed [3]. We have decided that a PCE is able to feed the TED of other PCEs; and using PCEP because it is the protocol designed for communicating them.

We can summarize I3TFP operation in a few steps:

1. I3TFP capability discovery. This makes possible for the local PCE to start I3TFP negotiation with other PCEs supporting interdomain TE information exchange.
2. TE parameter negotiation. To achieve an agreement about the TE information that is going to be exchanged.
3. TE information exchange. This operation is essential for our proposal. In this step, interdomain TED-to-TED feedback is performed and every involved PCE gets an amplified view of the global interdomain system.

For each one of these steps, the definition of new objects and TLVs (Type, Length, Value) are needed. One can see these modifications and additions in the next paragraphs.

4.2. Discovering I3TFP capabilities

Before a given local PCE starts I3TFP operation, it needs to discover the I3TFP capabilities of its surrounding foreign PCE. This way it can be sure that they know the I3TFP messages it sends. The PCE capabilities are usually spread together with the PCE discovery messages of the PCE discovery mechanisms. Specifically, they are defined in CAP-FLAGS TLV [22], [32], [33] and SVT-CAP-FLAGS sub-TLV [22], that has the same format and meaning. These TLVs have a set of bits reserved for future capabilities definition and we have redefined one of them (bit number nine) to advertise I3TFP support, as can be shown in Table 2.

Table 2. Definition of I3TFP capability.

Bit 9 of CAP-FLAGS and SVT-CAP-FLAG	Meaning after redefinition
0	The PCE does not support I3TFP.
1	The PCE supports I3TFP.

Therefore, when the intra and interdomain PCE discovery process takes effect, the PCE will have a database of available peer PCEs and a set of capabilities of each one; among other, it will know whether a PCE support I3TFP operation or not. It will allow the PCE to establish I3TFP communication with the selected peer PCE when desired.

4.3. TE parameters negotiation

Since some aspects of PCEP operation are agreed during the establishment of the PCEP session, it is necessary to introduce some new features in the PCE messages in charge of this establishment in order to negotiate the use of I3TFP. A given PCE implementing I3TFP could be configured to share not all the available TE parameters but only some of them. This fact facilitates the configuration of dissimilar exchange policies depending, for example, on the partner in the I3TFP conversation. For that reason, it is desirable that the PCEs involved in a PCEP session can achieve an agreement about the common TE parameters they are willing to use during their cooperation.

The OPEN message [11], and particularly the OPEN object it contains, carries some optional TLVs indicating the aspirations of the initiating PCE for the PCEP session. The PCE in the other side of the communication has to send a response telling whether it agrees to the conditions expressed in these TLVs or not. If it does not agree, it will send a PCERR message [11] back to the initiating PCE, including an OPEN object containing the same TLVs but including the alternate set of characteristics it is willing to accept for the PCEP session (Fig. 6).

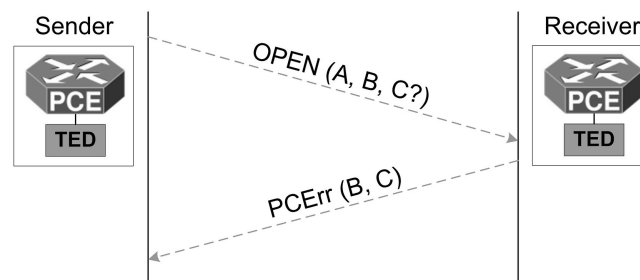


Fig. 6. Example of a generic PCEP session negotiation

The OPEN message is subject to extensions by means of the definition of new TLV containing features that should be negotiated before the establishment of the PCEP session. We have designed a new TLV, called I3TFP-NEG (Fig. 7), which allows a PCE to bargain the TE parameters to be exchanged while the PCEP session is alive.

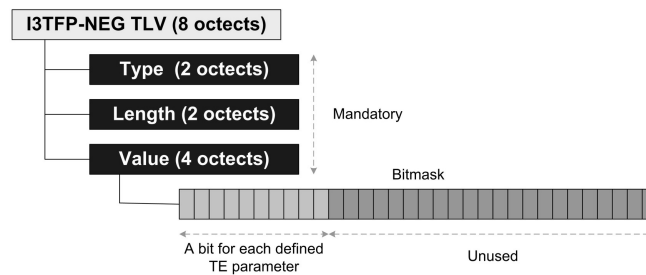


Fig. 7. Format of I3TFP-NEG TLV

The value field is a four-octet bitmask where only the first ten less significant bits are used and the rest are reserved for future TE parameter definitions. Each one of the defined ones references a given TE parameter in the order shown in Table 1. The PCE that starts the PCEP session should set to 1 the bits corresponding to those TE parameters it wants to exchange during the PCE session; and it should set the rest to 0. The negotiation follows the usual rules. Upon receiving unacceptable PCEP session characteristics, the receiving peer PCEP has to include an OPEN object in the PCERR message to propose the TE parameters they agree to share.

When the PCEP session features have been agreed and it has been established, both PCEs know the PCE parameters they are going to share to feed the TED of each other.

4.4. Exchanging abstract TE information

Once the peers have agreed the set of aggregated parameters they are going to share, we have to provide them with a set of PCEP extensions to make feasible this exchange. The NOTIFICATION message (PCNTf) of PCEP [8] allows the advertisement of relevant events that should be known by both elements. This message and particularly its NOTIFICATION object can be extended by designing new optional TLVs with new meaning. We have defined new Notification Type (NT) and Notification Value (NV) for the header of the NOTIFICATION object (Table 3) to denote the existence of relevant changes in the interdomain resources.

Table 3. Definition of new NT and NV values.

NT	NV	Meaning
3	1	Aggregated resources update. Some changes in the resources of the interdomain system have happened. One or more I3TFP-UPD is attached.

Following the NOTIFICATION object header, the sender has to attach a set of optional TLVs to be interpreted by the receiver. In the case of an I3TFP advertisement, we have designed a new TLV called I3TFP-UPD (Fig. 8) that will encapsulate completely the I3TFP sub protocol. A NOTIFICATION object could have more than one occurrence of the I3TFP-UPD TLV, but at least one.

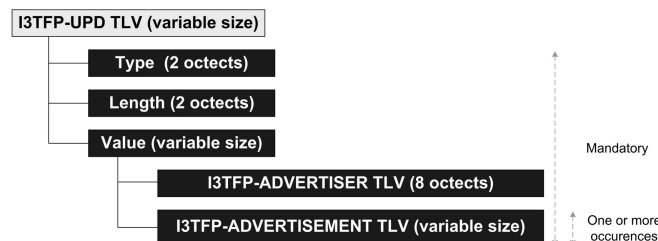


Fig. 8. Format of I3TFP-UPD TLV

The aim of the I3TFP-UPD TLV is to feed the RICUBE of the PCE. Since the aggregated information is stored in the RICUBE as three values (adjacent domain, target domain, TE parameter), this new TLV has to provide the receiver with this information. First, it announces the advertiser PCE by means of the I3TFP-ADVERTISER TLV (Fig. 9). This fixed-size TLV has a four-octet field that will contain the AS (Autonomous System) number of the domain the advertiser PCE belongs to. This way, the receiver is able to set it as the adjacent domain in the RI-CUBE.

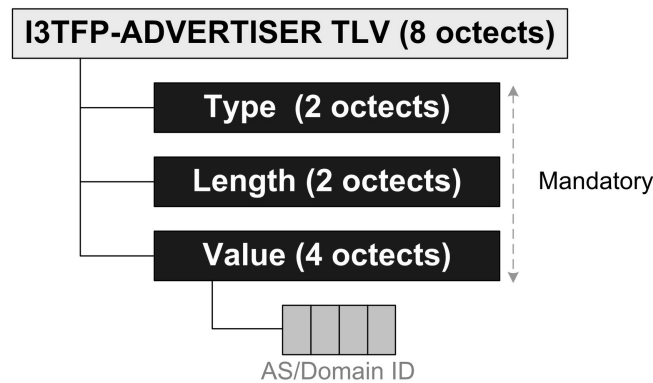


Fig. 9. Format of I3TFP-ADVERTISER TLV

Apart from the advertiser domain, the I3TFP-UPD TLV carries information related to the aggregated resources that have changed (and caused the advertisement) through the I3TFP-ADVERTISEMENT TLV (Fig 10). I3TFP-UPD TLV can include more than an I3TFP-ADVERTISEMENT TLV so it is a variable-size TLV. As the announcer of the update is known, each I3TFP-ADVERTISEMENT has to provide the receiver PCE with the other still unknown two values: the target domain and the TE parameter value. Note that the sender is always the same for every I3TFP-ADVERTISEMENT TLV, but the target domain could be different because the PCE in the adjacent domain will supply the local one with information related to every reachable I3TFP-supporting domains it knows, either directly or indirectly.

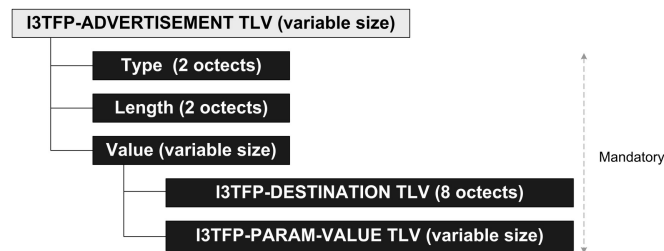


Fig. 10. Format of I3TFP-ADVERTISEMENT TLV

The target domain of a specific advertisement is provided to the receiver PCE via the I3TFP-DESTINATION TLV (Fig. 11). The format of this TLV is the same that I3TFP-ADVERTISER. Nevertheless, the meaning is different because it refers only to a particular advertisement and not to the overall I3TFP update announce.

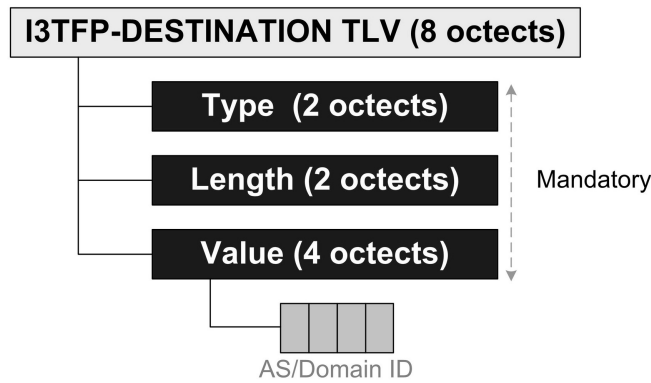


Fig. 11. Format of I3TFP-DESTINATION TLV

Finishing, the last TLV we have designed is I3TFP-PARAM-VALUE (Fig. 12). It gives the receiver PCE the updated aggregated value for a given TE parameter from the point of view of the sender PCE. In this fashion, the sender tells the receiver the expected state, in terms of resources and capabilities, to reach a given target domain if it is the selected peer PCE to compute an interdomain LSP towards this target domain. In other words, it is advising the receiver how is the situation of the area of the network to be crossed from itself to the target domain. This data is enough for the local PCE to predict, taking in mind the constraints to be applied to the LSP computation, whether it is a good idea to choose the sender as a cooperating PCE or not. However, it is insufficient to infer neither the contribution of each domain to this piece of information nor the number of domains to traverse.

Since I3TFP-PARAM-VALUE can include information related to each one of the ten selected parameters (Table 1), it is a variable-size TLV. The receiver PCE should analyse the Value field. It is composed by two fields: a four-octet field called Sub-type and a variable field called Value whose size depends on the subtype specified in Sub-Type. Table 4 shows the values we have defined for Sub-type field and the corresponding size for Value. The size of Value depends on the nature of the TE parameters it stores. Anyway, in order to fulfil the requirements of the PCEP TLV format expressed in [8], the size the Value field will be the lower four-octet multiple that is able to store the aggregated information.

Sub-types 1-10 refer to parameters 1-10 as shown in Table 1. Hence, for example, if a PCE receives an advertisement containing an I3TFP-PARAM-VALUE TLV with Sub-type 4, it will notice that the sender is advertising the new aggregated value for the TE parameter 4 (Maximum available bandwidth for the LSP) and therefore, the Value field size will be 32 octets (an array of 8 four-octet values).

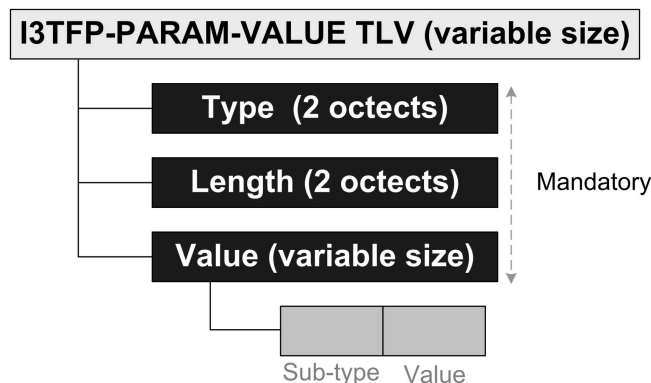


Fig. 12. Format of I3TFP-PARAM-VALUE TLV

Moreover, there are some situations where it is desirable the exchange of new values of all the parameters related to a specific target domain, for example, when turning on a new I3TFP-enhanced PCE or after a network reconfiguration. For this reason, we have defined an additional I3TFP-PARAM-VALUE Sub-type=0 (ALL) to accomplish the transmission of new values for all the defined TE parameters avoiding the excessive overhead of using an I3TFP-ADVERTISEMENT TLV per parameter. This I3TFP-PARAM-VALUE sub-type can also be used when the number of TE parameters that needs an update makes the bandwidth consumption greater than the use of a single ALL advertisement.

Table 4. I3TFP-PARAM-VALUE sub-types

Sub-type	Value length	Format of the aggregated value
0 (ALL)	96 octets	The same format of sub-types 1 to 10 put together in ascending order.
1 (MAXBW)	4 octets	32 bits IEEE 754.
2 (MAXRBW)	4 octets	32 bits IEEE 754.
3 (UBW)	32 octets	8x32 bits IEEE 754. One for each one of the eight existing priority levels.
4 (MAXBWLSP)	32 octets	8x32 bits IEEE 754. One for each one of the eight existing priority levels.
5 (LLPT)	4 octets	Values 0x01, 0x02, 0x04, 0x08, 0x10 and 0x20, depending on the aggregated local link protection
6 (MINBWLSP)	4 octets	32 bits IEEE 754.
7 (MTU)	4 octets	A number $[0-2^{16}]$.
8 (MPLSTE)	4 octets	Boolean meaning.
9 (GMPLS)	4 octets	Boolean meaning.
10 (SC)	4 octets	Values 1, 2, 3, 4, 51, 100, 151 and 200, depending on the aggregated switching capability.
≥ 11	Reserved	Reserved

At this point, I3TFP-enhanced PCEs have the PCEP extensions they need, I3TFP, to distribute their particular view of the interdomain system, its resources and its possibilities. After the PCEP session negotiation, they know the traffic engineering parameters they agree to share; now, they have the mechanisms to make effective this transmission.

After the PCEP session takes effect, I3TFP-UPD TLVs are sent asynchronously by both PCEs when required (Fig. 13). There is not an agreed updating time. Some thresholds maintained independently by each PCE determine the virtual updating time. When a change in the local resources or in resources learned from others (RI-CUBE change) exceeds the threshold defined locally for a particular TE parameter, that PCE will send an advertisement, using this proposal, to update the TED of the PCEs in the foreign adjacent domains. The process to set the correct value for these thresholds is outside the scope of this work, although it should be based on measurements and statistics related to the variation of the resources in each specific domain, in order to dampen the oscillations in too-changing domains.

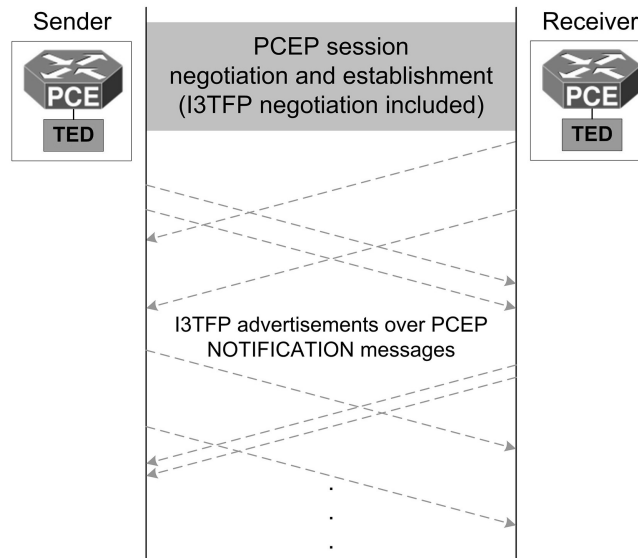


Fig. 13. Asynchronous I3TFP advertisements

4.5. I3TFP granularity and updating time effects

There are some aspects to analyse in relation to the proposed set of PCEP extensions. As we said in the last section, I3TFP allows multiple granularities when advertising changes in the resources of the interdomain system: an I3TFP-UPD per advertisement, a single I3TFP-UPD containing all the pending advertisements or a single I3TFP-UPD containing a single advertisement using I3TFP-PARAM-VALUE TLV with Sub-type=ALL. Fig. 14 shows the differences of using these types of advertisements to announce one hundred changes in the resources of interdomain TE parameters. In the chart, one can see the effect of the overhead. When spreading the one hundred changes, the bandwidth consumption can vary between ≈ 1 KB and ≈ 3 KB.

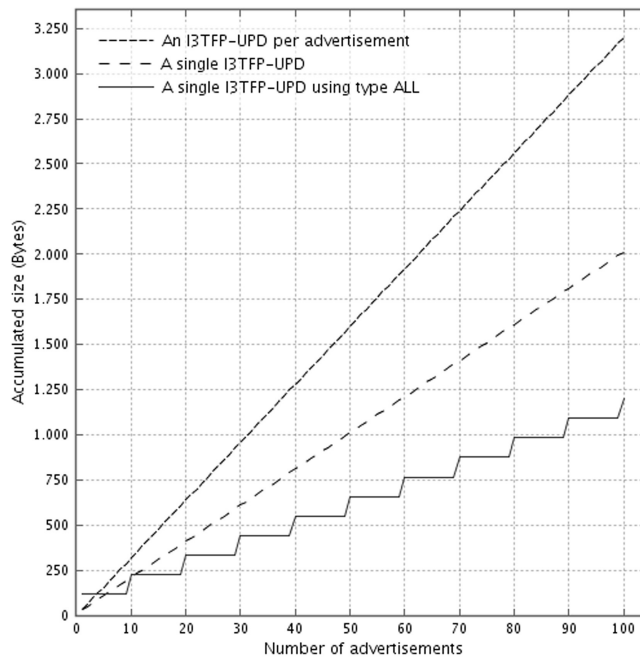


Fig. 14. Effects of I3TFP granularity on the bandwidth consumption

However, this is an extreme case where there are not mixes in the way the advertisements are sent. It is an explanatory example. In a running network, the sender PCE should consider whether to use a given mechanism or another, in order to save bandwidth by reducing the number of advertising messages. It will not use always the same and fixed kind of I3TFP advertisement. In addition, not all the types of advertisement can be used in all the situations since it depends on some factors as, for example, the target domain, the advertisements quantity or the TE parameter they refer to.

Other factor that affects the bandwidth consumption is the updating time. As we said before, there is not a negotiated updating time in the I3TFP operation. Instead, the I3TFP advertisements are sent when needed. Nonetheless, the frequency of these advertisements, which depends on the configured value of some thresholds, has an impact in the bandwidth required by the transmission of I3TFP messages. Fig. 15 shows an example where one can see the overall bandwidth used by I3TFP operation after an hour. There, the sender announces a single I3TFP-PARAM-VALUE TLV per I3TFP-UPD TLV (the worst case) following a different updating time: 30, 60 and 120 seconds. The result is that the updating time is inversely proportional to the bandwidth consumed by I3TFP messages.

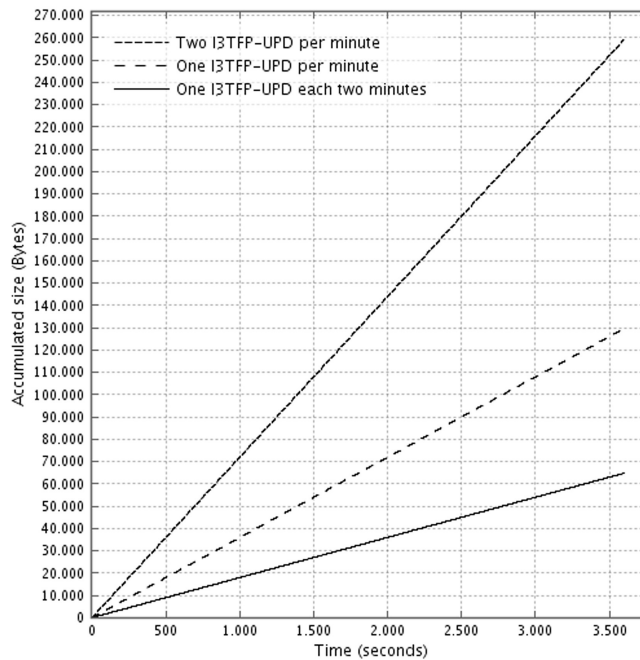


Fig. 15. Effects of I3TFP updating time on the bandwidth consumption

In general, as shown in Fig. 16, the updating time and the I3TFP granularity used are related to the effectiveness of the PCE selection mechanism and the bandwidth consumption. The PCE selection mechanism we propose requires the existence of updated TE information inside the RI-CUBE; using a lower updating time will bring that information more updated and, consequently, the PCE selection effectiveness will be higher. However, this will be translated into an increase of the bandwidth consumption. For that reason, it will be necessary to find out a balance between these two factors. Anyway, independently of the selected updating time, a PCE implementing I3TFP should choose the adequate I3TFP granularity to group the advertisements it needs to send, in order to decrease the needed bandwidth.

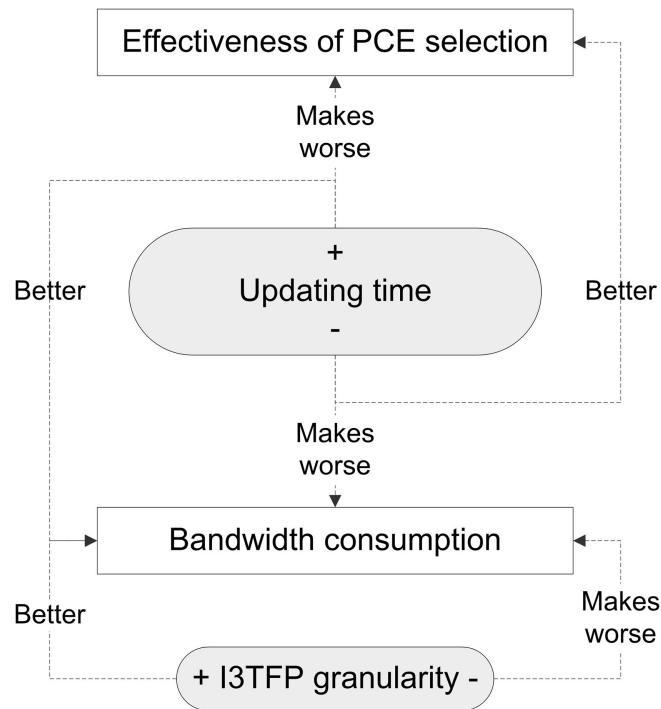


Fig. 16. Relationship between granularity, bandwidth updating time and effectiveness

5. Conclusions and future work

In this work, we suggest I3TFP, a PCEP sub protocol composed by PCEP extensions that allow a PCE to distribute its aggregated TE information to others. All these extensions have been carefully designed so that they are transparent for the usual operation of PCEP. Thus, we have supplied the community with a mechanism to exchange TE information, safely, through the aggregation of meaningful, common and comparable TE information among all the domains. Having the TE information in its TED makes possible for an interdomain-capable PCE to pre-process the constraints belonging to a given LSP computation request, to choose the best foreign PCE in a more accurate way.

In the near future, we have planned to advance in this proposal by designing fast and useful pre-processing algorithm to exploit the aggregated TE information that is available now.

Acknowledgements

This research work is sponsored in part by the regional Education, Science and Technology Ministry belonging to the Extremadura (Spain) Regional Government, through the AGILA-2 project numbered as PR1A06145; and by the national Industry, Tourism and Commerce Ministry belonging to Spanish National Government, through MESEAS project numbered as FIT-350301-2007-14.

References

- [1] D. Awduche, A. Chiu, A. Elwalid, I. Widjaja, X. Xiao. Overview and Principles of Internet Traffic Engineering. IETF RFC 3272. May, 2002.
- [2] E. Crawley, R. Nair, B. Rajagopalan, H. Sandick. A Framework for QoS-based Routing in the Internet. IETF RFC 2386. August, 1998.

- [3] A. Farrel, J. P. Vasseur, J. Ash. A Path Computation Element (PCE)-Based Architecture. IETF RFC 4655. August, 2006.
- [4] E. Rosen, A. Viswanathan, R. Callon. Multiprotocol Label Switching Architecture. IETF RFC 3031. January, 2001.
- [5] E. Mannie. Generalized Multi-Protocol Label Switching (GMPLS) Architecture. IETF RFC 3945. October, 2004.
- [6] E. Oki, I. Inoue, K. Shiimoto. Path computation element (PCE)-based traffic engineering in MPLS and GMPLS networks. IEEE Sarnoff Symposium 2007. April 30 2007-May 2 2007. pp. 1-5. Digital Object Identifier: 10.1109/SARNOF.2007.4567400.
- [7] T. Saad, J. Israr, S. Sivabalan, H.T. Mouftah. An Evaluation for PCE Selection Schemes for Inter-Domain Path Computation. 9th International Conference on Transparent Optical Networks, 2007. ICTON '07. Volume 3, 1-5. July 2007 pp. 187-187. DOI 10.1109/ICTON.2007.4296276.
- [8] M. Domínguez-Dorado, José-Luis González-Sánchez, J. Domingo-Pascual, J. Carmona-Murillo. RI-CUBE: dotando al PCE de información abstracta de ingeniería de tráfico interdominio. Proceedings of the V Iberoamerican Conference on Telematics (CITA'09). pp. 14-21. ISBN 978-84-613-2679-2. Gijón (SPAIN), May, 2009.
- [9] J. Ash, J.L. Le Roux. Path Computation Element (PCE) Communication Protocol Generic Requirements. IETF RFC 4657. September, 2006.
- [10] J.L. Le Roux. Path Computation Element Communication Protocol (PCECP) Specific Requirements for Inter-Area MPLS and GMPLS Traffic Engineering. IETF RFC 4927. June, 2007.
- [11] J. P. Vasseur, J. L. Le Roux. Path Computation Element (PCE) communication Protocol (PCEP). IETF RFC 5440. March, 2009.
- [12] Y. Lee, G. Bernstein. Alternative Approaches to Traffic Engineering Database Creation and Maintenance for Path Computation Elements. IETF Draft draft-lee-pce-ted-alternatives-02.txt. May, 2009. Work in progress.
- [13] D. Awduche, L. Berger, D. Gan, T. Li, V. Srinivasan, G. Swallow. RSVP-TE: Extensions to RSVP for LSP Tunnels. IETF RFC 3209. December 2001.
- [14] M. Yannuzzi, X. Masip-Bruin, O. Bonaventure. Open issues in interdomain routing: a survey. IEEE Network. Volume 9, Issue 6. November-December, 2005.
- [15] C. Pelsser, S. Uhlig, O. Bonaventure. On the difficulty of establishing interdomain LSPs. Proceedings of the IEEE Workshop on IP Operations and Management. October 2004.
- [16] R. Zhang, J.P. Vasseur. MPLS Inter-Autonomous System (AS) Traffic Engineering (TE) Requirements. IETF RFC 4216. November, 2005.
- [17] J.L. Le Roux, J.P. Vasseur, J. Boyle. Requirements for Inter-Area MPLS Traffic Engineering. IETF RFC 4105. June, 2005.
- [18] K. Kumaki, T. Murai. BGP protocol extensions for Path Computation Element (PCE) Discovery in a BGP/MPLS IP-VPN. IETF Draft draft-kumaki-pce-bgp-disco-attribute-03.txt. March, 2009. Work in progress.
- [19] Vijayanand, Somen Bhattacharya, Prasanna Kumar. BGP protocol extensions for PCE Discovery across Autonomous Systems. IETF Draft draft-vijay-somen-pce-disco-proto-bgp-04.txt. July, 2007. Expired I-D.
- [20] M. Boucadair, P. Morand. Path Computation Service discovery via Border Gateway Protocol. May, 2005. IETF Draft draft-boucadair-pce-discovery-01.txt. Expired I-D.
- [21] M. Boucadair, P. Morand. A Solution for providing inter-AS MPLS-based QoS tunnels. IETF Draft draft-boucadair-pce-interas-01.txt. May 2005. Expired I-D.
- [22] M. Domínguez-Dorado, José-Luis González-Sánchez, J. Domingo-Pascual. PILEP: a contribution to PCE-based interdomain path computation. Proceedings of the 13th International Telecommunications Network Strategy and Planning Symposium (NETWORKS'08). pp. 1-7. ISBN 978-963-8111-68-5. Budapest (HUNGARY), October, 2008.
- [23] J.L. Le Roux. Requirements for Path Computation Element (PCE) Discovery. IETF RFC 4674. October, 2006.
- [24] K. Kompella, D. Yeung. Traffic Engineering (TE) Extensions to OSPF Version 2. IETF RFC 3630. September, 2004.
- [25] H. Smit, T. Li. Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE). IETF RFC 3784. June, 2004.
- [26] P. Srisuresh, P. Joseph. OSPF-xTE: Experimental Extension to OSPF for Traffic Engineering. IETF RFC 4973. July, 2007.
- [27] K. Kompella, Y. Rekhter. OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS). IETF RFC 4203. October, 2005.

- [28] K. Kompella, Y. Rekhter. Intermediate System to Intermediate System (IS-IS) Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS). IETF RFC 4205. October, 2005.
- [29] J.P. Vasseur, S. Previdi. Definition of an IS-IS Link Attribute Sub-TLV. IETF RFC 5029. September, 2007.
- [30] J.P. Vasseur, J.L. Le Roux. IGP Routing Protocol Extensions for Discovery of Traffic Engineering Node Capabilities. IETF RFC 5073. December, 2007.
- [31] A. Lindem, N. Shen, J.P. Vasseur, R. Aggarwal, S. Shaffer. Extensions to OSPF for Advertising Optional Router Capabilities. IETF RFC 4970. July, 2007.
- [32] J.L. Le Roux, J.P. Vasseur, Y. Ikejiri, R. Zhang. OSPF Protocol Extensions for Path Computation Element (PCE) Discovery. IETF RFC 5088. January, 2008.
- [33] J.L. Le Roux, J.P. Vasseur, Y. Ikejiri, R. Zhang. IS-IS Protocol Extensions for Path Computation Element (PCE) Discovery. IETF RFC 5089. January, 2008.
- [34] M. Domínguez-Dorado, José-Luis González-Sánchez, J. Domingo-Pascual. Descubrimiento de PCE inter-AS: una aportación a la computación de LSP en sistemas multidominio. Libro de actas de las VII Jornadas de Ingeniería Telemática (JITEL'08). Págs. 80-86. ISBN 978-84-612-5474-3. Alcalá de Henares (ESPAÑA), Septiembre de 2008.