ETSI GR NGP 014 V1.1.1 (2019-10)

GROUP REPORT

Next Generation Protocols (NGP); Preferred Path Routing (PPR) for Next Generation Protocols

Disclaimer

Reference
DGR/NGP-0014

Keywords
mobility, QoS, routing support

ETSI

650 Route des Lucioles
F-06921 Sophia Antipolis Cedex - FRANCE

Tel.: +33 4 92 94 42 00 Fax: +33 4 93 65 47 16

Siret N° 348 623 562 00017 - NAF 742 C
Association à but non lucratif enregistrée à la
Sous-Préfecture de Grasse (06) N° 7803/88

Important notice

The present document can be downloaded from:
http://www.etsi.org/standards-search

The present document may be made available in electronic versions and/or in print. The content of any electronic and/or print versions of the present document shall not be modified without the prior written authorization of ETSI. In case of any existing or perceived difference in contents between such versions and/or in print, the prevailing version of an ETSI deliverable is the one made publicly available in PDF format at www.etsi.org/deliver.

Users of the present document should be aware that the document may be subject to revision or change of status. Information on the current status of this and other ETSI documents is available at https://portal.etsi.org/TB/ETSIDeliverableStatus.aspx

If you find errors in the present document, please send your comment to one of the following services:
https://portal.etsi.org/People/CommiteeSupportStaff.aspx

Copyright Notification

Contents

Intellectual Property Rights

Essential patents

IPRs essential or potentially essential to normative deliverables may have been declared to ETSI. The information pertaining to these essential IPRs, if any, is publicly available for **ETSI members and non-members**, and can be found in ETSI SR 000 314: *"Intellectual Property Rights (IPRs); Essential, or potentially Essential, IPRs notified to ETSI in respect of ETSI standards"*, which is available from the ETSI Secretariat. Latest updates are available on the ETSI Web server (https://ipr.etsi.org/).

Pursuant to the ETSI IPR Policy, no investigation, including IPR searches, has been carried out by ETSI. No guarantee can be given as to the existence of other IPRs not referenced in ETSI SR 000 314 (or the updates on the ETSI Web server) which are, or may be, or may become, essential to the present document.

Trademarks

The present document may include trademarks and/or tradenames which are asserted and/or registered by their owners. ETSI claims no ownership of these except for any which are indicated as being the property of ETSI, and conveys no right to use or reproduce any trademark and/or tradename. Mention of those trademarks in the present document does not constitute an endorsement by ETSI of products, services or organizations associated with those trademarks.

Foreword

This Group Report (GR) has been produced by ETSI Industry Specification Group (ISG) Next Generation Protocols (NGP).

Modal verbs terminology

In the present document "**should**", "**should not**", "**may**", "**need not**", "**will**", "**will not**", "**can**" and "**cannot**" are to be interpreted as described in clause 3.2 of the ETSI Drafting Rules (Verbal forms for the expression of provisions).

"**must**" and "**must not**" are **NOT** allowed in ETSI deliverables except when used in direct citation.

Executive summary

The present document focuses on using Preferred Path Routing (PPR) for next generation architectures towards various access, transport and DC networks and beyond. The basic concept of PPR consists of a novel path routing paradigm for various data planes that allows to provide dynamic traffic engineering optionally with resource reservation along the path. PPR can be deployed in a way that supports current architectures, while also enabling more optimal future architectures. The work aims to examine and propose recommendations to improve and simplify the network infrastructure to support optimized source routing aligned with SDN/NFV infrastructure natively by adopting PPR. In addition, the present document may require the development of new protocols and/ or modification of existing protocols.

Introduction

The present document provides recommendations toward new protocols and/or modification of existing ones in the context of PPR.

1 Scope

The present document provides brief overview of existing routing mechanisms, traffic engineering and proposes next generation source routing which can support multiple and extensible data planes with hard SLA guarantees with dynamic resource reservations along the path. It explores new TE framework for high-precision transport networks by signalling simple linear paths as well as graph structures to provide compact and yet provide better scalability of overall paths in the network.

2 References

2.1 Normative references

Normative references are not applicable in the present document.

2.2 Informative references

References are either specific (identified by date of publication and/or edition number or version number) or non-specific. For specific references, only the cited version applies. For non-specific references, the latest version of the referenced document (including any amendments) applies.

NOTE: While any hyperlinks included in this clause were valid at the time of publication, ETSI cannot guarantee their long term validity.

The following referenced documents are not necessary for the application of the present document but they assist the user with regard to a particular subject area.

[i.1] ISO/IEC 10589:2002 (Second edition): "Intermediate system to intermediate system intra-domain-routing routine information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)".

[i.2] IETF RFC 2328: "OSPF Version 2", Moy, J., April 1998.

NOTE: Available at https://www.rfc-editor.org/info/rfc2328.

[i.3] IETF RFC 5340: "OSPF for IPv6", Coltun, R., Ferguson, D., Moy, J., and A. Lindem, July 2008.

NOTE: Available at https://www.rfc-editor.org/info/rfc5340.

[i.4] IEEE Global Communications Conference (GLOBECOM): "The Segment Routing Architecture", C. Filsfils, N. K. Nainar, C. Pignataro, J. C. Cardona, P. Francois, 2015, San Diego, CA, 2015.

[i.5] draft-ietf-spring-segment-routing-mpls-13: "Segment Routing with MPLS data plane", A. Bashandy, C. Filsfils, S. Previdi, B. Decraene, S. Litkowski, R. Shakir, (work in progress), IETF, April 2018.

[i.6] draft-ietf-6man-segment-routing-header-12: "IPv6 Segment Routing Header (SRH)", S. Previdi, C. Filsfils, J. Leddy, S. Matsushima, D. Voyer, (work in progress), IETF, April 2018.

[i.7] IETF RFC 5440: "Path Computation Element (PCE) Communication Protocol (PCEP)", J.P. Vasseur, J.L. Le Roux, March 2009.

NOTE: Available at https://www.rfc-editor.org/info/rfc7752.

[i.8] draft-ietf-isis-segment-routing-msd-10"Signaling MSD (Maximum SID Depth) using IS-IS", J. Tantsura, U. Chunduri, S. Aldrin, L. Ginsberg, (work in progress), IETF, April 2018.

[i.9] IETF RFC 5714: "IP Fast Reroute Framework", Shand, M. and S. Bryant, January 2010.

[i.10] IETF RFC 7490: "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, April 2015.

NOTE: Available at http://www.rfc-editor.org/info/rfc7490.

[i.11] IEEE Global Communications Conference (GLOBECOM): "Preferred Path Routing - A Next-Generation Routing Framework Beyond Segment Routing", U. Chunduri, A. Clemm, R. Li, 2018, Abu Dhabi, UAE, December 2018.

[i.12] IEEE Hi-Precision Networks (HIPNET): "Preferred Path Routing (PPR) Graphs Beyond signalling of paths to Networks", T. Eckert, Y. Qu, U. Chunduri, 2018, Rome, Italy, 2018.

[i.13] IETF RFC 2025: "The Simple Public-Key GSS-API Mechanism (SPKM)".

[i.14] IETF RFC 3209: "RSVP-TE: Extensions to RSVP for LSP Tunnels".

[i.15] IETF RFC 7810: "IS-IS Traffic Engineering (TE) Metric Extensions".

[i.16] IETF RFC 7471: "OSPF Traffic Engineering (TE) Metric Extensions".

[i.17] IETF RFC 8402: "Segment Routing Architecture".

[i.18] draft-ietf-rtgwg-segment-routing-ti-lfa-01: "Topology Independent Fast Reroute using Segment Routing".

[i.19] IETF RFC 8300: "Network Service Header (NSH)".

NOTE: Available at https://tools.ietf.org/html/rfc8300.

3 Definition of terms, symbols and abbreviations

3.1 Terms

Void.

3.2 Symbols

Void.

3.3 Abbreviations

For the purposes of the present document, the following abbreviations apply:

3GPP	3rd Generation Partnership Project
5G	Fifth Generation Mobile Networks
BGP	Border Gateway Protocol
CPU	Central Processing Unit
CSR	Cell Site Router
DB	DataBase
DC	Data Centre
DCI	Data Centre Interconnect
EH	IPv6 Extension Headers
FEC	Forwarding Equivalence Class
FIB	Forwarding Information Base
FRR	Fast ReRoute
GPRS	General Packet Radio Service
GTP	GPRS Tunneling Protocol
HW	Hardware
IGP	Interior Gateway Protocol
IP	Internet Protocol
IS-IS	Intermediate System to Intermediate System
LDP	Label Distribution Protocol

LFA	Loop Free Alternative
LSA	Link State Advertisement (OSPF)
LSP	Link State PDU (IS-IS)
LTE	Long Term Evolution
MPLS	Multi-Protocol Label Switching
MSD	Maximum SID Depths
MSDC	Massive Scale Data Centre
MTU	Maximum Transmission Unit
NGER	Next Generation Explicit Routing
NH	NextHop
NR	New Radio
OAM	Operations, Administration and Maintenance
OSPF	Open Shortest Path First
PCE	Path Computation Element
PDE	Path Description Element
PE	Provider Edge
PPG	Preferred Path Graph
PPR	Preferred Path Routing
PPR-ID	Preferred Path Route Identifier
PPR-TE	Preferred Path Route- Traffic Engineering
QFI	QoS Flow Identifier
RQI	Reflective QoS Indicator
RSVP	Resource Reservation Protocol
SDN	Software-Defined Networks
SID	Segment IDentifier
SLA	Service Level Agreement
SP	Service Provider
SPF	Shortest Path First
SR	Segment Routing
SRH	Segment Routing Header
TCP	Transmission Control Protocol
TE	Traffic Engineering
TLV	Type Length Value
UDP	User Datagram Protocol
UE	User Equipment
UPF	Use Plane Functionality (5G)
WI	Work Item

4 Background

4.1 Overview

Routing is a fundamental concept in packet networks. This clause provides background about routing technologies that are dominant today and points out certain shortcomings. This sets the stage for the introduction of PPR in the subsequent clause.

4.2 Shortest-path routing

Much of today's routing is based on the concept of Shortest Path Routing, based on the idea to always attempt to route packets along a path that is the "shortest", or of the least cost.

Research of SPF algorithm (invented by Dijkstra) started in 1970's and variations/modifications of this core algorithm is widely deployed with link state protocols or IGPs (OSPFv2 [i.2], IS-IS [i.1], OSPFv3 [i.3]). In IGPs, a directed graph is computed with flooded link state information (LSA/LSP DB) with links having configured weights/metrics. SPF Algorithm calculates a tree of shortest path from self to all other nodes in the network with candidate list of nodes kept sorted by weight. Shortest (best) value in the candidate and downloaded to the routing table with the computed immediate Next-Hop (NH). IP routing table only needs NH to each advertised prefix from all the nodes while LSA/LSP tree has all the paths. In the example below a shortest path from Rs is shown to a prefix advertised from Rd is shown with NH set to R11. Similar to Rd all Rs would compute NHs for all prefixes advertised from all the nodes in the network.

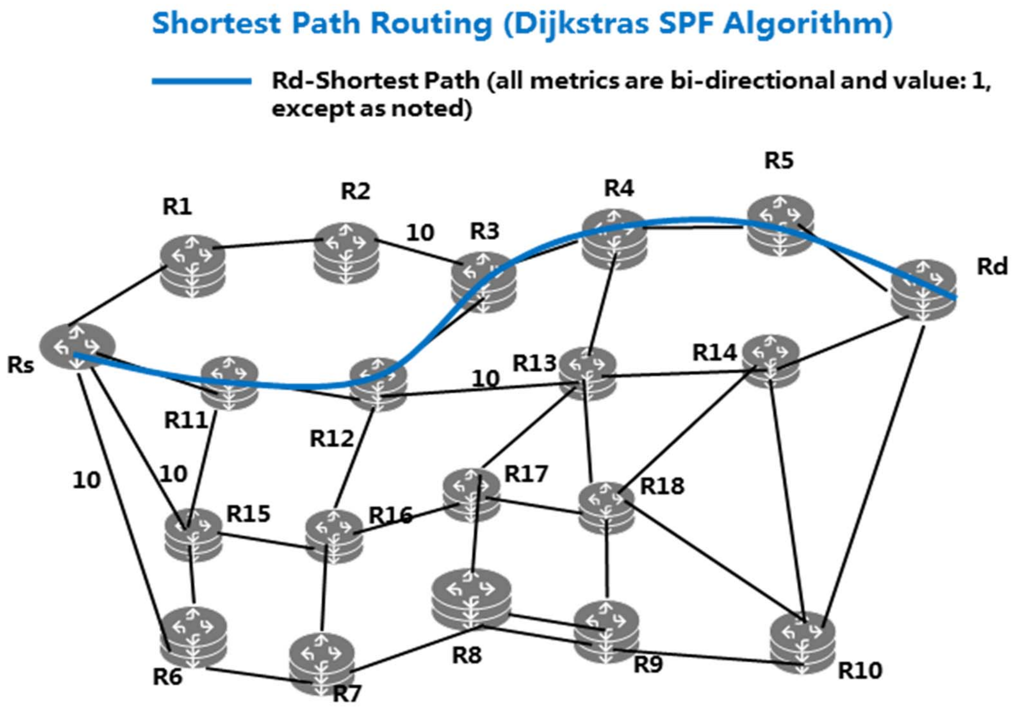

Figure 1: Concept of a shortest path

One drawback of shortest-path routing concerns that it is not always the shortest path that may be preferred, as there may be different cost metrics and other considerations (such as load balancing, ease of failover service levels, or robustness of path). As a result, other technology has begun to appear that allows to route on paths other than shortest paths. One such technology is Segment Routing (see clause 4.3).

4.3 Segment Routing

Segment Routing [i.4] is a novel source routing approach, which enables packet steering with a specified path in the packet itself. This is defined for MPLS (with a set of stacked labels) and IPv6 (path described as list of IPv6 addresses in SRHeader), and data planes called SR-MPLS [i.5] and SRv6 [i.6] respectively.

SR simplifies the Multi-Protocol Label Switching (MPLS) control plane by distributing Segment Identifiers (SIDs) for routing prefixes, which are constitute MPLS global labels into Interior Gateway Protocols. This allows source routing to be achieved by representing the network path with stacked SIDs on the data packet without any changes to the MPLS data plane. In addition to MPLS, as specified above, SR also introduces an IPv6 Extension Header (EH) for use with the IPv6 data plane, resulting in SRv6. In SRv6, a segment is encoded as an IPv6 address, with a new type of IPv6 Routing Header (EH) called SRH. A set of segments is encoded as an ordered list of IPv6 addresses in SRH to represent the path of the data packet.

Segments and source routes can be computed by a controller with knowledge of the network topology and subsequently provision the network with end-to-end SR paths. A controller could include e.g. a Path Computation Element (PCE) [i.7] or another type of SDN controller. Using a controller allows to perform different optimization and customizations to paths that take into account different constraints. This also obviates the need for traditional MPLS control plane protocols like LDP and RSVP, reducing the number of protocols that need to be deployed in a network.

To illustrate, Figure 2 depicts an example of an SR Path. To represent this path, a stack of 8 labels from node R15 to Rd is needed. In fact, more than 8 labels may be required to accommodate entropy labels, which become necessary for better load balancing of traffic across MPLS networks.

However there are some issues/drawbacks with SR:

1) The additional path overhead with complete path using SIDS on the data packet in various SR deployments may cause the following issues:

   - HW Capabilities: Not all nodes in the path can support the ability to push or read label stack needed Maximum SID Depth (MSD) [i.8] to satisfy user/operator requirements. Alternate paths which meet these user/operator requirements may not be available.

   - Line Rate: Potential performance issues in deployments, which use SRH data plane with the increased size of the SRH with 16 byte SIDs.

   - MTU: Larger SID stacks on the data packet can cause potential MTU/fragmentation issues.

   - Header Tax: Some deployments, such as 5G, require minimal packet overhead in order to conserve network resources. Carrying 40 or 50 octets of data in a packet with hundreds of octet of header would be an unacceptable use of available bandwidth.

2) Another limitation of SR concerns the fact that while it allows a data packet to steer through a custom path, by itself it cannot guarantee that proper QoS along the path needed. The ability to manage resource reservations or to provide traffic engineering attributes are not in SR's scope.

3) A more subtle issue concerns the ability to conduct performance measurements and collect traffic accounting statistics using SR-MPLS. Because labels on data packets refer only to individual path segments, attributing statistics of any particular packet to a path or flow is inhibited and difficult to perform efficiently.

4) SR cannot be applied to native IPv4/IPv6 data planes. While SR can be supported with MPLS without any changes in the data plane, use with IPv6 requires an SRH extension header, whose support requires hardware upgrades across the network. While SR is considered as a potential alternative for backhaul transport networks (like 5G), non-support for native IP data planes imposes a significant hurdle on SR adoption, as many cellular networks around the world still use native IPv4 and IPv6 data planes. As path steering capability is an essential component for network slicing in 5G backhaul transport, lack of this capability forces operators to upgrade the hardware for SRH support.

5) Last but not least SR also defines complex FRR approach with Topology Independent LFA (TI-LFA) [i.18]. Here, the post convergent backup path does not reflect the original SR path QoS characteristics. This is because alternative path is computed in a distributed fashion by the network nodes using LFA/RLFA [i.9] and [i.10] algorithms which can only give a loop free shortest path to the destination.

5 Preferred path routing concept and architecture

5.1 Overview

PPR is a new source routing paradigm where a prefix is signalled in a routing domain (control plane) along with a data plane identifier as well as path description on how the packets has to be forwarded when actual data traffic with the data plane identifier is seen. This builds on existing IGPs and fits well with SDN paradigm as the needed path can be crafted dynamically based on various inputs from a central entity.

5.2 PPR Core Concept

Traditionally routing in network is based on shortest path computations (through Interior Gateway Protocols or IGPs [i.1], [i.2] and [i.3]) for all prefixes in the network. As explained, Segment Routing allows to compute custom paths (other than shortest paths) that are subsequently represented by a sequence of segment identifiers in a packet, leading to another set of problems.

Preferred Path Routing (PPR) enables route computation based on a specific path described along with the prefix as opposed to shortest path towards the prefix. The key change that is required concerns how the next hop is computed for the prefix. Instead of using the next hop of the shortest path towards the destination, the next hop towards the next node in the path description is used. PPR is a novel architecture to signal explicit path and per-hop processing requirement and optionally including QoS or resources to be reserved along the path.

PPR is concerned with the creation of a routing path as specified in the PPR-Path which is advertised in IGPs along with a data plane identifier (PPR-ID). With this, any packet destined to the PPR-ID would use the PPR-Path instead of the IGP computed shortest path to the destination indicated by the PPR-ID. In other words, packets destined to the PPR-ID may use the PPR-Path instead of the IGP computed shortest path. This works as follows: IGP nodes process the PPR-Path. If an IGP node finds itself in the PPR-Path, it sets the next-hop towards the PPR-ID according to the PPR-Path.

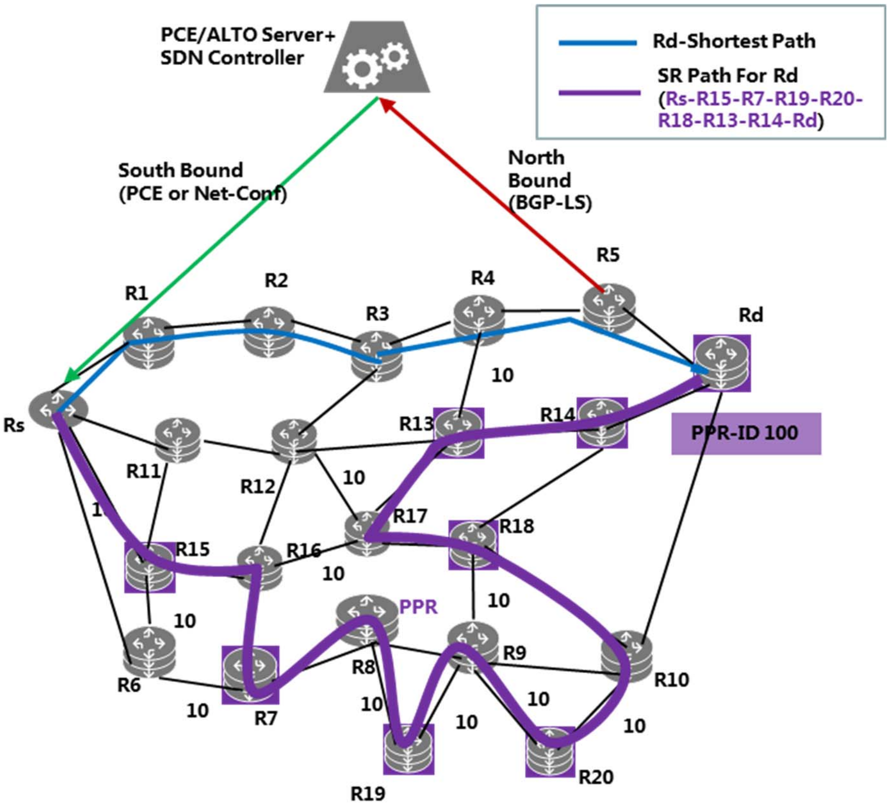

Figure 2: Shortest Path, Source Routed Path

Consider the network depicted earlier in Figure 2 to describe the operation of PPR. Node Rs is an ingress or head-end node, while node Rd serves as egress or as another head-end node. Assume the bi-directional IGP link metric for all the links connecting any two node to be of value 1 except from some links with value 10, as shown explicitly. Rs may be configured to receive TE or explicit source routed path information from a central entity (PCE or Controller). The received path comprises PPR information. A PPR is identified using a PPR-ID, which can also relate to sources that are attached to node Rs: traffic from those sources may have to use a specific PPR-ID. (It is also possible to have a PPR provisioned locally for non-TE needs, e.g. for purposes of FRR or to chain services.) The PPR path information is encoded as an ordered list of PPR-Path Description Elements (PDEs) from source to a destination node in the network. The PPR-PDE information represents both topological and non-topological segments and specifies the actual path towards a Forwarding Equivalence Class (FEC) or Prefix by Rd.

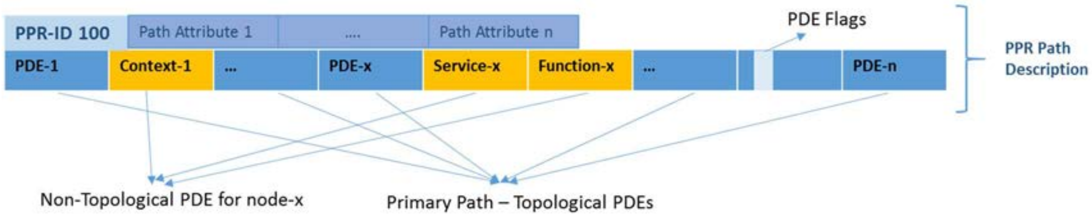

Figure 3: PPR Path Structure

Considering the same example, the SR path Rs-R15-R7-R19-R20-R18-R13-R14-Rd can be attached with a PPR-ID, say 100. Once the path and PPR-ID are signalled in an underlying IGP as a PPR, only nodes that find themselves in the path description have to act on this path. For example, after completing its shortest path computation as usual, R15 finds that its node information is encoded as PPR-PDE in the path. As a result, it adds an entry to its Forwarding Information Base (FIB) with PPR-ID 100 as the incoming label (assuming the data plane type in PPR TLV is MPLS) and sets the NH as the shortest path NH towards the next PPR-PDE (R7), which in this case is the link towards R16. If instead R15 had added a shortest path route entry in the FIB for Rd, it would have added by setting NH as link towards R11 (shortest path metric to reach Rd). This process continues on every node as represented in the PPR path description.

Inter-Area Scenarios:

- The above can be extended inter-area scenarios. The below diagram (Figure 4) represents one such scenario. In this 2 IS-IS levels are used each having separate north bound and south bound communication end points with PCE/SDN controller. It is expected PPR path for each level is computed and given to the ingress nodes Rs and Rd for L1 and L2 respectively. Node Rd is acting as an L1/L2 router and some of the prefixes in L2 are leaked to L1 (including Rb). Now in L1 area, the path advertised by Rs (shown in purple) would be id for prefix Rb, hosted by L1/L2 router Rd. At Rd when it receives the PPR for Rb in L2, the same is advertised in L2 area (in orange), in this example it happened to be a strict path Rd-Ra-Ra15-Ra16-Ra17-Ra13-Ra14-Rb. It is important to note the L1/L2 router functionality of separate path advertisement in respective levels (no leaking of path information).

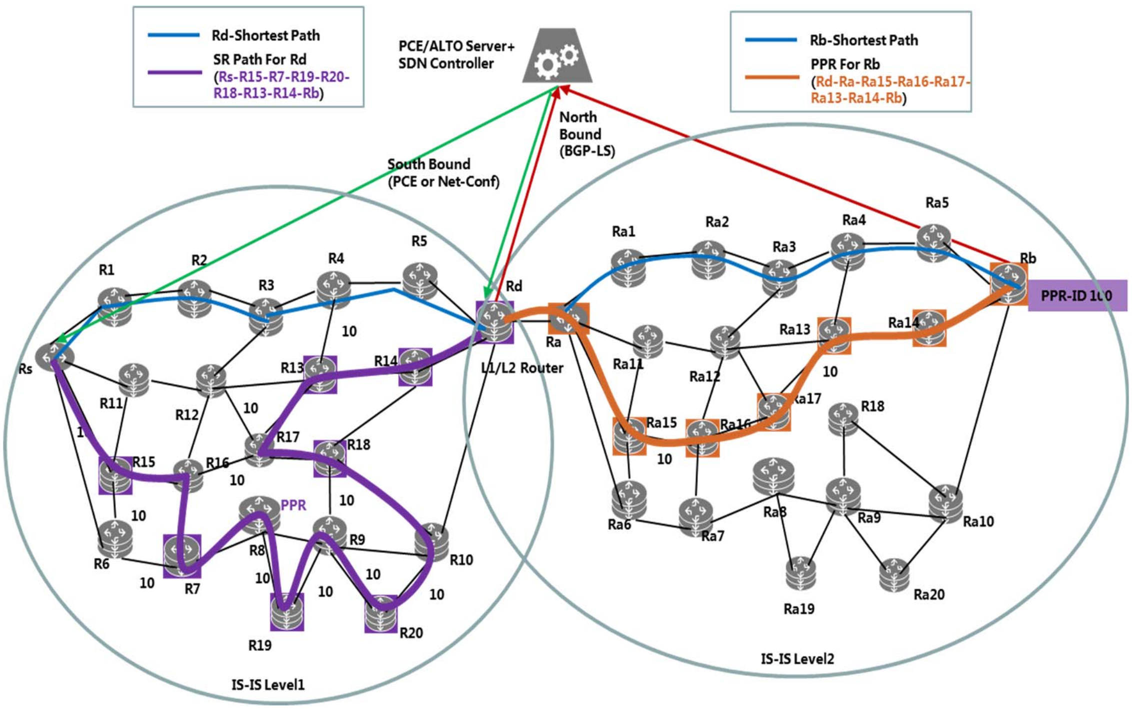

Figure 4: PPR with Inter-Area Scenario (IS-IS Example)

5.3 Loose and Strict paths

There are two variants according to which paths can be specified: loose and strict.

In case of a strict path, every node along the path is defined and aware of the PPR-ID. This means that the PPR-ID itself is sufficient for forwarding decisions and is the only label that needs to be carried.

In case of a loose path, some nodes along the path are specified and aware of the PPR-ID. In that case, there are intermediate path segments on which nodes are not aware of the PPR-ID; forwarding decisions are then based on the next node on the path that can be reached. In this case, the Segment ID defining the next node on the path is added to the packet (in addition to the PPR-ID), which is popped as the next node on the path is reached.

The example in Figure 2 describes a loose path as not every node in the path from Rs to Rd is specified; for example, for "…R15-R7…" portion of the path there is another node R16 (in addition to R6) over which R7 could be reached. The path type (loose or strict) is explicitly indicated in the PPR-ID description. R15 acts on this flag and in the case of a loose path programs the local hardware with two labels/SIDs, using PPR-ID 100 as bottom and node SID of R7 as top label. Intermediate nodes like R16 do not need to be aware of PPR and the fact that data packets are being transported along a PPR path. They just forward the packet based on the top label, in this case to R7. However, if the path described were a strict path, in an MPLS data plane the actual data packet would require only a single label, i.e. PPR-ID 100.

5.4 Services along with the path

As shown in Figure 3 some of the services can be encoded as non-topological Path Description Elements (PDEs) and can be part of the overall path. These services would be applied at the respective nodes along the path. In Figure 3, PDE-1, PDE-x, PDE-n are topological PDEs of the particular data plane. For SR-MPLS and SRv6 data planes these are simply SIDs. When the data packet with PPR-ID 100 is delivered to node-1, the packet is delivered with context-1. Similarly on node-x, service-x1 is applied and function-x1 is executed. These service and functions need to be pre-provisioned on the particular nodes and can be advertised in IGPs. These should be known to the central entity/controller along with Link State Database of IGP that is used in the underlying network.

The above gives the basic and light weight service chaining capability with PPR without incurring any additional overhead on the data packet. However this is limited to fixed functions/services for a particular path and all data packets using the path will be applied with these services. Flow levels exclusions using the same path or differentiated services that need to be applied with in a flow cannot be supported with this mechanism and one has to resort to full-blown NSH/SFC IETF RFC 8300 [i.19] data plane for the same.

5.5 Compatibility with SDN and PCE Architecture

While a PPR can be defined by operator from any node in the network by statically provisioning and signalling the same, PPR is most suitable for a centralized SDN paradigm. In the centralized paradigm complete topology information from the underlying network including dynamic TE information from the links as well as business logic to set up a TE path with certain TE parameters viz., x Bandwidth, y total path latency and z allowable jitter range required for certain traffic.

5.6 Extensibility of Source Routing: Native IP Data Planes

Another advantage of PPR is the ability to provide source routing and path steering capabilities to legacy IP networks without having to change hardware or even upgrade the data planes.

Figure 5 shows how PPR works with native IPv4 and IPv6 data planes. As shown in the figure, all links metrics are set to value 2 except the links from Rs to R6, and from R8 and R12, which have a link metric of value of 1o. The shortest path to reach Rd from Rs is via R1-R2-R3-R4-R5. In IPv4, by signalling a PPR path Rs-R6-R7-R8-R12-R13-R14-Rd with a PPR-ID set to an IPv4 address 10.10.10.1, any IPv4 packet with destination IP address set to 10.10.10.1 would be steered along the path. Similarly in IPv6, the same PPR path will be signalled with IPv6 address 2001::1. Note only Strict-PPR is applicable to these data planes.

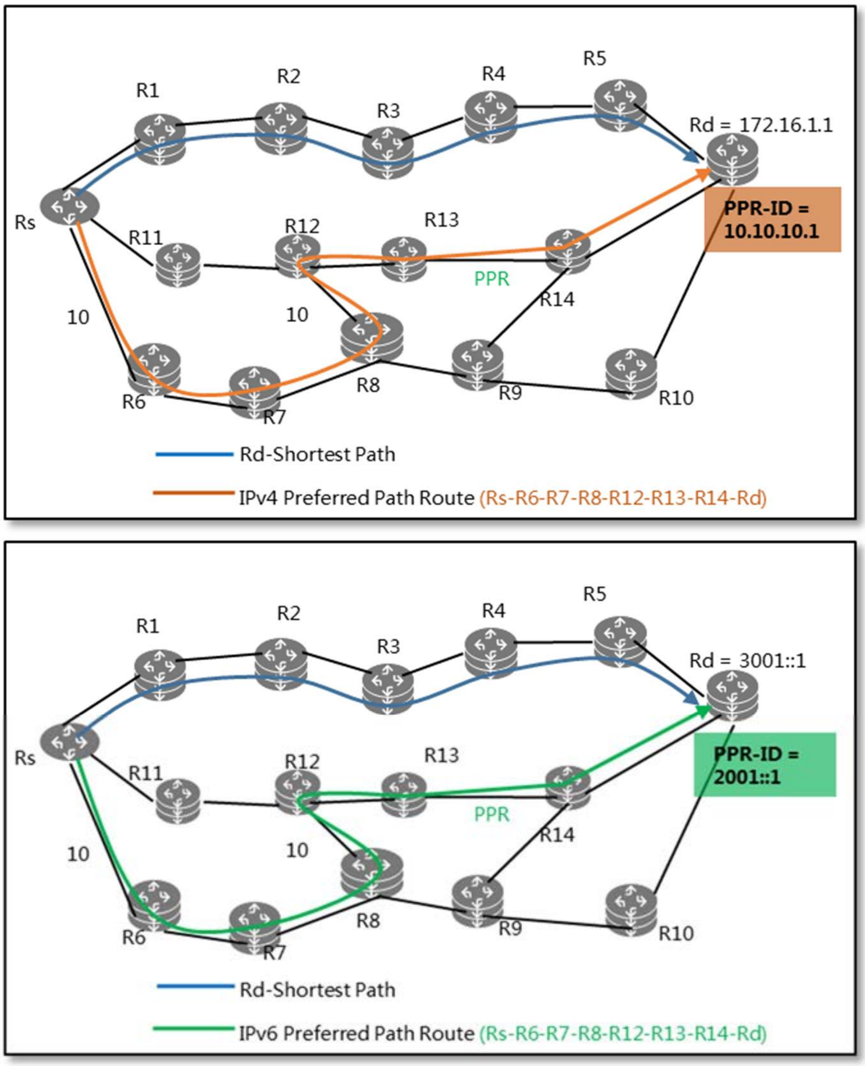

Figure 5: PPR for native IP (IPv4, IPv6) data planes

Data plane path identifier, PPR-ID enables data plane extensibility as the type of the data plane can be indicated as shown in Figure 6.

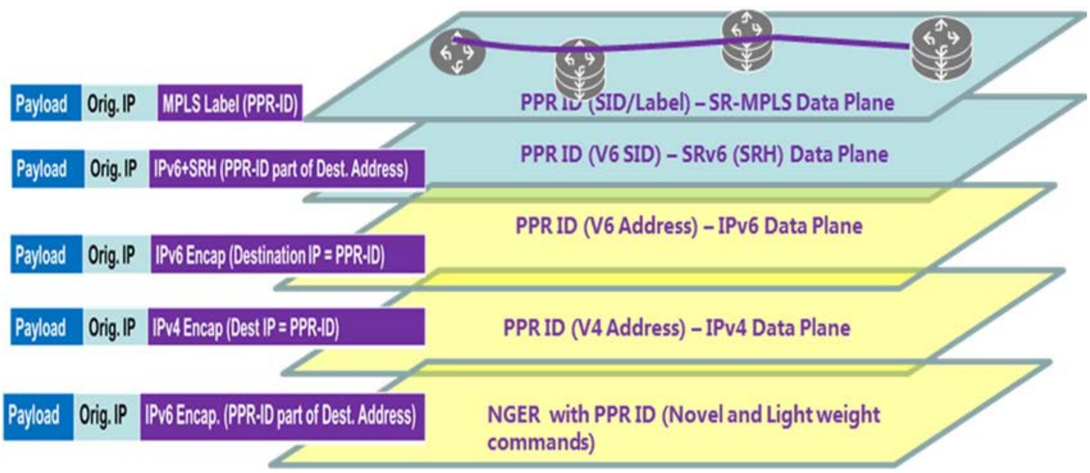

Figure 6: Extensible data planes with PPR

NGER: This is a new data plane abbreviated as Next-Generation Explicit routing where network programming can be done with no additional overhead on the packet other than the encapsulated IPv6 Header. Further details would be documented a separate work item.

5.7 PPR compatibility with SR and Optimized SR Data Planes

PPR is fully backward compatible with SR as PDEs can be extensible and particular data plane identifiers can be expressed to describe the path and in SR case PDEs can contain the SR SIDs (topological like nodal and adjacency SIDs or non-topological SR SIDs).

One of the key benefits PPR offers for SR data planes (SR-MPLS, SRv6) is providing the same benefits (e.g. of source routing based on a predefined path) with an optimized data plane with at most one or two labels on the packet for strict and loose cases respectively (as specified in clause 5.3).

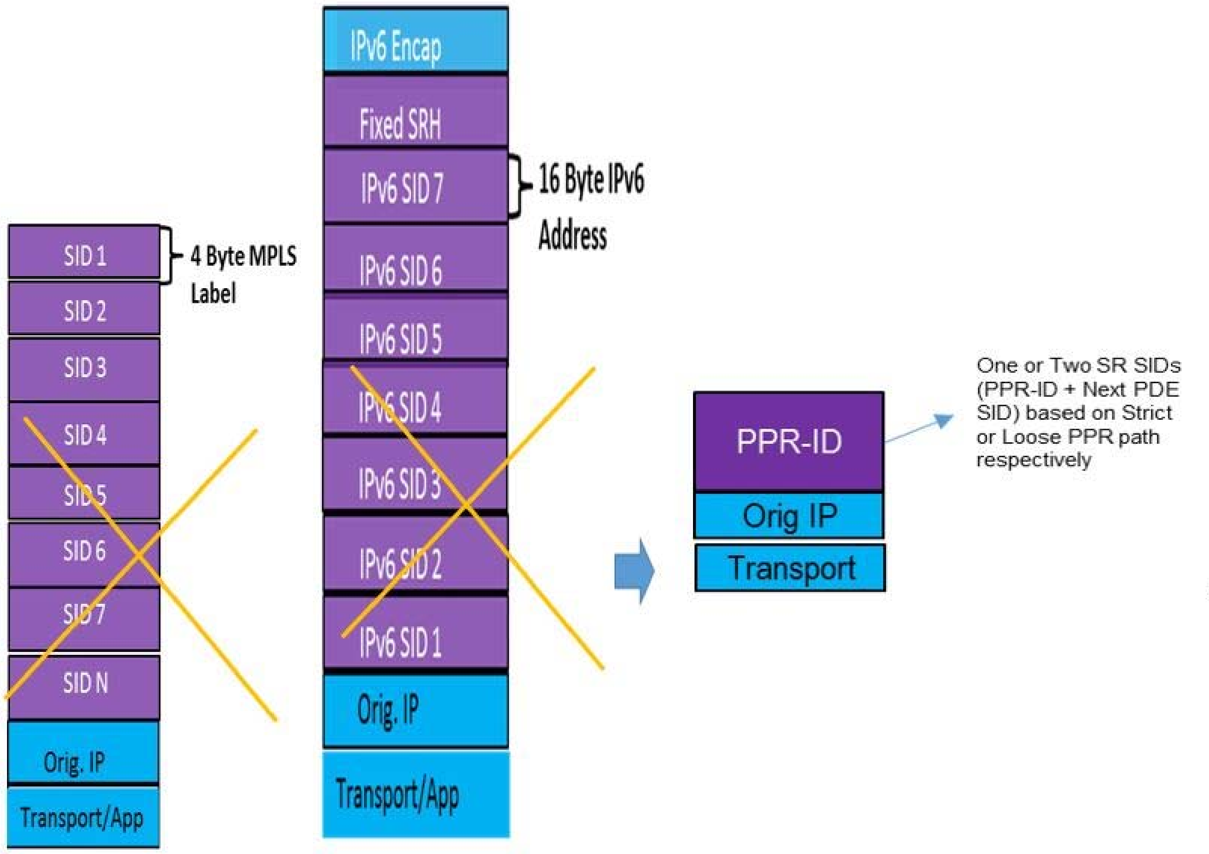

Figure 7: Simplified SR data plane (SR-MPLS, SRv6) with PPR

6 PPR and Resources for the Path

In addition to determining the nodes to traverse, there may be other aspects that need to be set up for a path. Most notably, this concerns the allocation and reservation of resources along the path in order to help ensure the service levels, i.e. the Quality of Service that is delivered across the path, will be acceptable for the traffic routed across the path.

RSVP (IETF RFC 2025 [i.13]) allows out of band signalling along a specified path for resource reservations. This is done by sending PATH/RESV message with flow spec/filter spec. IETF RFC 3209 [i.14], builds on RSVP protocol and defines new objects, modifies existing objects for MPLS LSP establishment. This is not considered as dynamic and requires provisioning along the path with out of band signalling. Also it relies on periodic refreshes for state synchronization between neighbours.

Segment Routing [i.5] and [i.6] enables packet steering with a specified path in the packet itself. This is defined for MPLS (with stacked labels) and IPv6 (path described as list of IPv6 addresses in SRHeader) data planes. Generally a controller computes the path and installs the same at ingress nodes with path description and as per local policy data flows are mapped to these paths. While this allows packet steering on a specified path, it does not have any notion of QoS or resources reserved along the path.

The determination of which resources to allocate and reserve on nodes across the path, like the determination of the path itself, can in many cases be made by a controller. Accordingly, PPR includes extensions that allow to manage those reservations, in addition to the path itself.

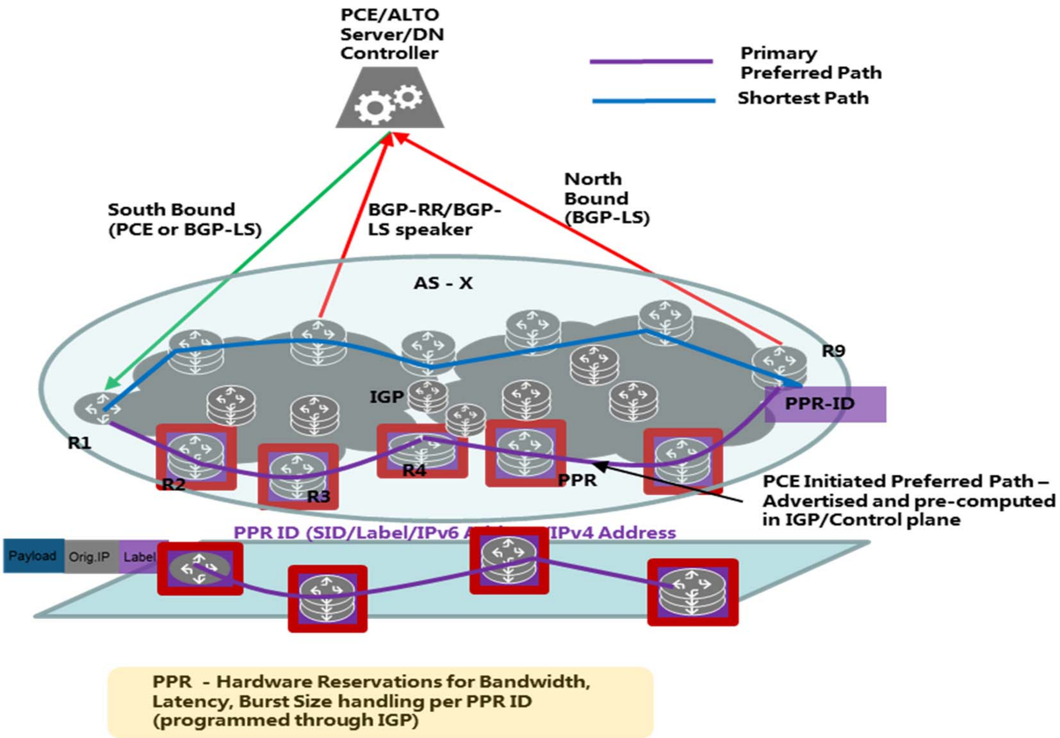

Figure 8: PPR with Resource Reservations along the path

Key aspect of the solution concerns with specifying the resources to be reserved along the preferred path, through path attributes TLVs. Reservations are expressed in terms of required resources (bandwidth), traffic characteristics (burst size), and service level parameters (expected maximum latency at each hop) based on the capabilities of each node and link along the path. The second part of the solution is providing mechanism to indicate the status of the reservations requested i.e. if these have been honoured by individual node/links in the path. This is done by defining a new TLV/Sub-TLV in respective IGPS. Another aspect is additional node level TLVs and extensions to IETF RFC 7810 [i.15] and IETF RFC 7471 [i.16] to provide accounting/usage statistics that have to be maintained at each node per preferred path. All the above is specified for IS-IS/OSPFv2/OSPFv3 protocols.

7 Scalability & PPR graphs

7.1 Overview

With PPR paths as described in earlier clauses, a separate PPR-ID is required for every possible path, even if one path is just a subset of another path. To provide PPR-Paths from N possible source nodes to one destination node, N PPR-IDs are necessary, and N PPR-Paths have to be signalled across the IGP. To create full-mesh connectivity via PPR-Paths between N nodes, o(N^2) PPR-Paths and PPR-IDs are necessary.

Even if in a larger network, only a smaller percentage of nodes want to use PPR-Paths, this can create a potential scale issue in terms of the number of forwarding entries that are required (one per PPR-ID), along with the associated number of topological records that need to be exchanged. In other words, impacts include the load on IGP flooding machinery to disseminate the PPR-Paths into the entire IGP area and additional FIB/label/forwarding entries for PPR-IDs.

This clause describes the various scaling aspects of PPR paths and different graph structures. Specifically, ways in which the number of PPR-IDs can be reduced are discussed, which occurs in part by generalizing paths into tree structures.

7.2 S & D Bits in PPR path structures

In PPR as defined in earlier clauses, a distinct linear path structure and PPR-ID is used for every path from a source node to a destination node. However, with a bit on path element "source" and a bit for "destination" - the same path ID/PPR-ID can be used to represent multiple paths if some of the nodes are also sources and terminating on the same destination node, as described in Figure 9.

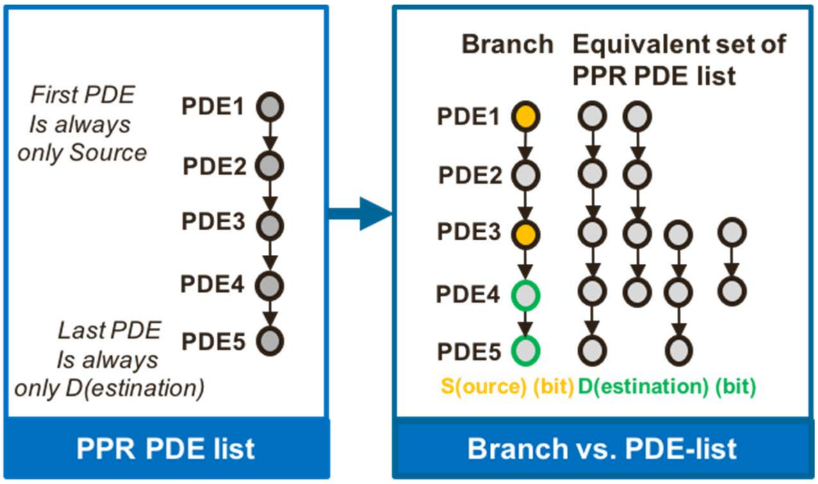

Figure 9: PPR Path Structure with S and D bits

In the above a "PPR PDE list" example, a linear path list of 5 nodes (called PDE - "Path Description Element") are described where PDE1 is the source/ingress-point and PDE5 is the destination/egress point of the path. On the right example "Branch vs. PDE-list", a Branch as introduced by this invention is shown, where every PDE can have a S(ource) and/or D(estination) bit. PDE1 and PDE3 have the Source bit, PDE4 and PDE5 the Destination bit. This Branch structure is equivalent to the set of 4 PPR-PDE lists shown on the right: PDE1->PDE5, PDE1->PDE5, PDE3->PDE4, PDE3->PDE5. This reduces the amount of information that needs to be sent across the IGP and that needs to be processed by each node.

In addition, whereas the 4 PPR PDE lists would have required each a unique PPR-ID (and the resulting forwarding entries created), the Branch requires only 2 PPR-IDs: one for both paths terminating in PDE4, and one for both paths terminating in PDE5.

Example 2 in Figure 10. On the left hand, 3 PPR Paths are shown. On the right hand side, this is replaced by a new PPR-Path with a Branch where Rs/R6 have Source bit set and R9/Rd have Destination bit set. On the left hand side, 6 PPR-IDs are required, on the right hand side, only 2 PPR-ID are required. Note that on the right hand side, also the path Rs- > R9 is permitted by the single PPR-Path.

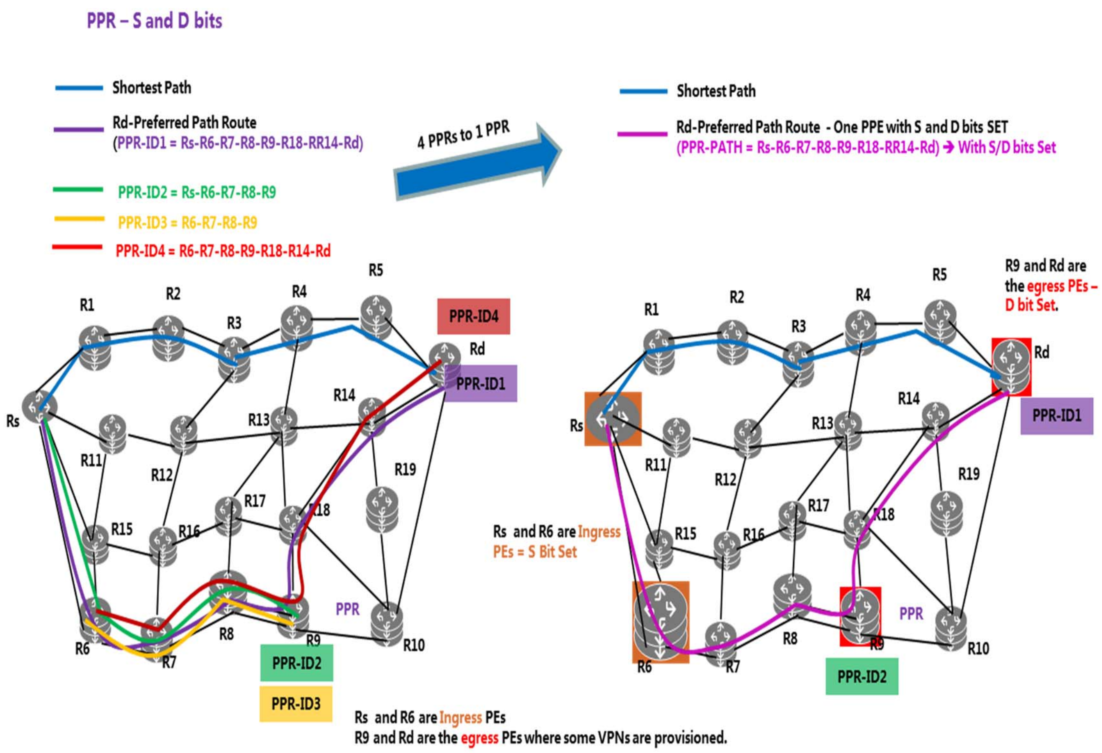

Figure 10: PPR with S and D bits for better path scalability

7.3 PPR Graph Structure

For optimization, the linear point-to-point PPR paths described earlier (in which every node has at most one predecessor and at most one successor) can be generalized into graphs. The simplest PPR graph structure is that of a tree graph where the root of the tree is the only destination in the graph and all edges point towards that root. Leave nodes of the tree will be sources and transit nodes may be sources. This type of graph is called a PPR Tree. PPR Trees are a compact representation of paths from multiple sources to a single destination that can be used whenever the paths from these sources to that destination converge on some node in the network and do not diverge afterwards.

When the desired paths from a set of nodes do not converge, multiple PPR trees may be required to represent the paths. In Figure 11, paths from source PE1, PE2, PE3 and PE4 to destination PE5 are represented as two PPR Trees. This choice of paths could for example be necessary because the aggregate traffic from all source PE's would be too much for link L1 into PE5, so for PE3 and PE4, another path via L2 needs to be used. In result, P1 needs to have different forwarding entries and therefore different PPR-IDs towards PE5. Likewise, P2 will need to establish forwarding entries for both PPR-IDs, because it is a transit node on both PPR Trees, whereas all other nodes only establish one or no PPR-ID forwarding entries for these two trees.

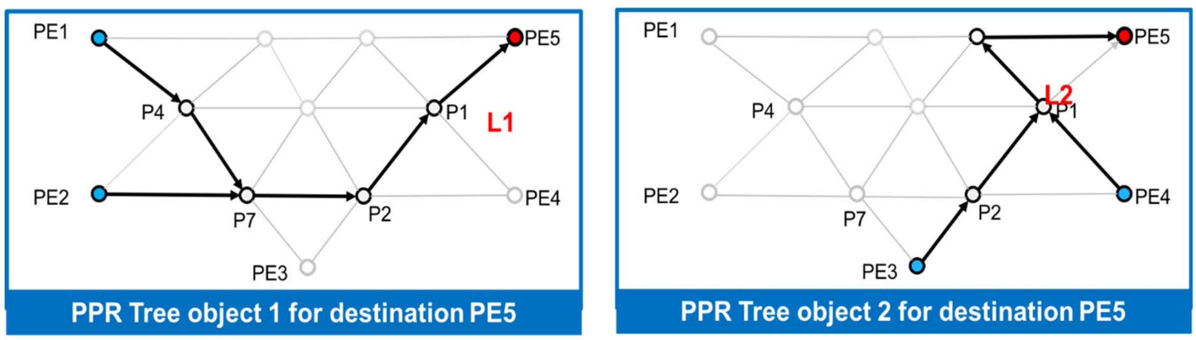

Figure 11: Different PPR Tree objects for same destination

For the linear encoding of PPR Graph structures, see Figure 12 below: A PPR Graph can be constructed from one or more PPR Branches. Branches are stitched together by using the same PDE (path description element) in two branches, e.g. C1 in Branch 3 and Branch 1. The resulting graph represents the set of all possible PPR paths from any PDE with S(ource) bit set to any PDE with D(estination) bit set.

To simplify parsing of branches, only the last PDE of a branch can be stitched to another branch. In result, any PDE can only be a non-last PDE in one Branch but last PDE in more than one branch.

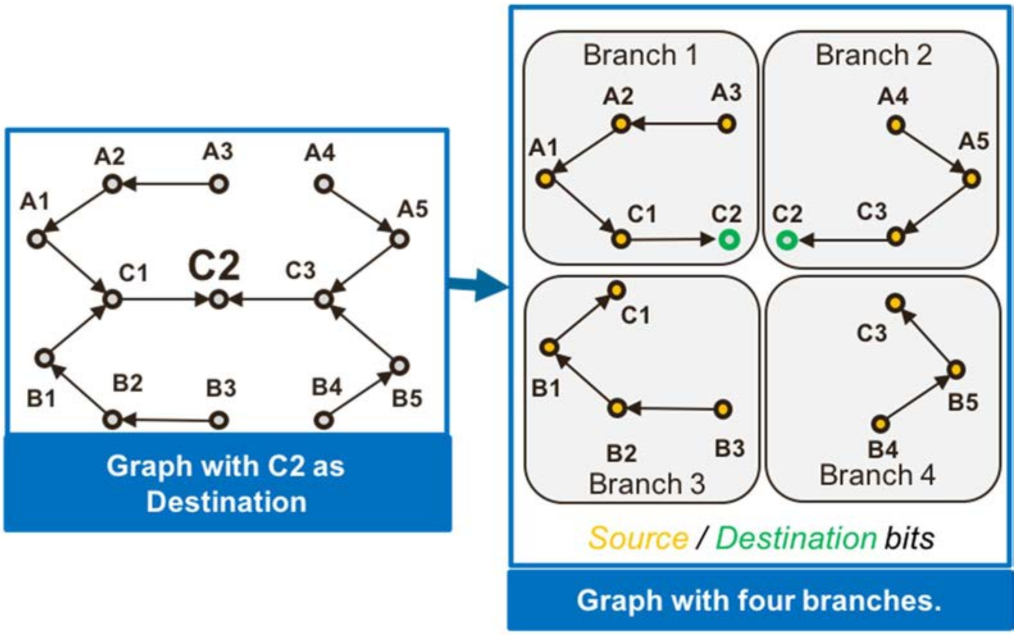

Figure 12: PPR Graph Structures

The right example is equivalent to 12 pre-existing PPR-ID Paths: Any source path element (PDE) to destination path element (PDE) C2.The key new encoding element of PPR Graphs over prior simple PPR Paths is the existence of multiple Branches in the PPR Graph description. Each branch-ID sub-TLV is followed by ordered sequence of PDEs.

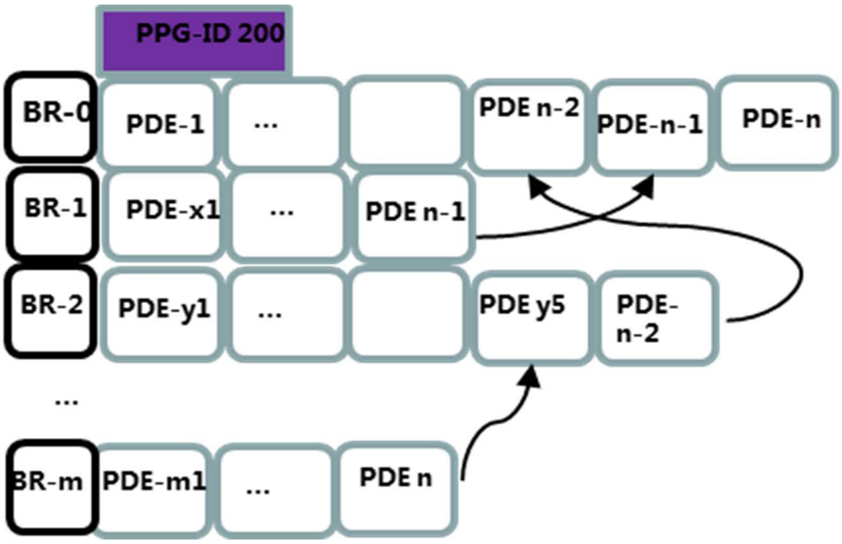

Figure 13: PPR Graph Structure Possible Encoding

A new PPG-ID sub-TLV can be defined for this purpose. It will have another sub-tlv code in PPR-TLV and rest of the structure is similar to PPR-ID structure. This can be part of the forwarding when the whole tree need to be used for sending and receiving the data. Individual branches can be formed from the graphs from each S to D nodes. And for the same PPR-ID at node with D bit set should be used in the forwarding.

Currently, the present document describes four types of PPR Graphs. The type of the PPR Graph can be encoded in the PPR "Type" TLV. These four graph types define unambiguous paths from one or more sources to one or more destinations.

PPR-Paths have one source and one destination. They can be segmented (like the other PPG Graph types) as explained below in "segmentation". They may not need to be encoded as a separate PPR graph type but instead are simply an instance of a PPR Tree. Explicitly calling them out as a separate PPR graph type may be helpful for diagnostics and to be able to verify that the graph complies with the constraint of having only one source.

PPR-Trees: One or more sources, one destination.

PPR-(unidirectional-) Forests: Two or more destinations, one or more sources.

PPR-Bidir-Forests: Bidirectional forests. This is the only graph type considered, where the connections between nodes are bidirectional.

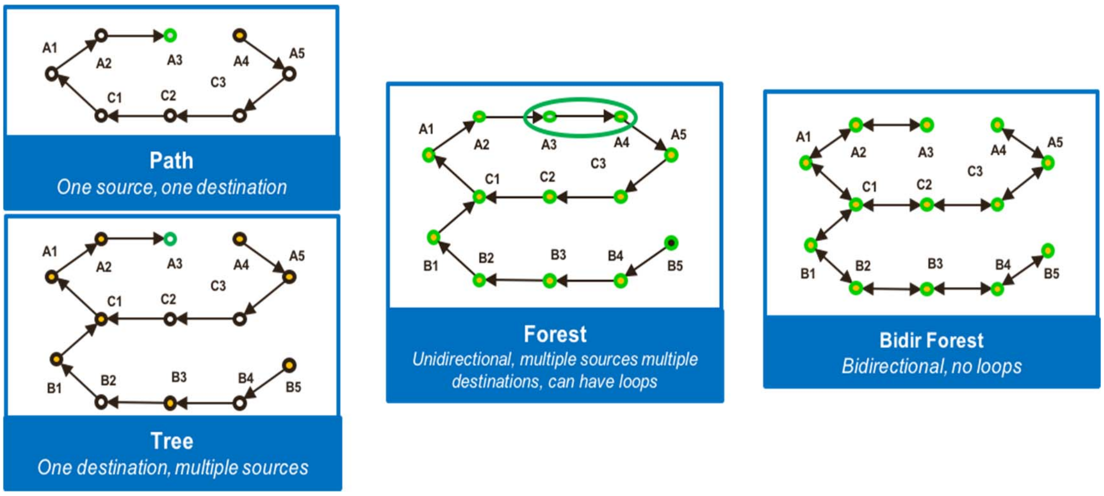

Figure 14: PPR Graphs (Trees, Forest, Bidir Forest)

Nodes processing a PPR Graph will ignore the whole graph if it does not comply with the Rules described for its Graph Type. Graphs that would describe ambiguous paths between sources/receivers are not permitted because they would leave it up to the receiving nodes to choose one of those paths and that would make then not well predictable by an external control system. Such graphs might be added in the future if there is demand for them though. Such ambiguous graphs could for example be used to represent the equivalent of IGP topologies, but this is outside the scope of the present document.

7.4 PPR Trees

The principles of PPR Trees where already introduced above. This clause describes the exact rules for PPR Trees.

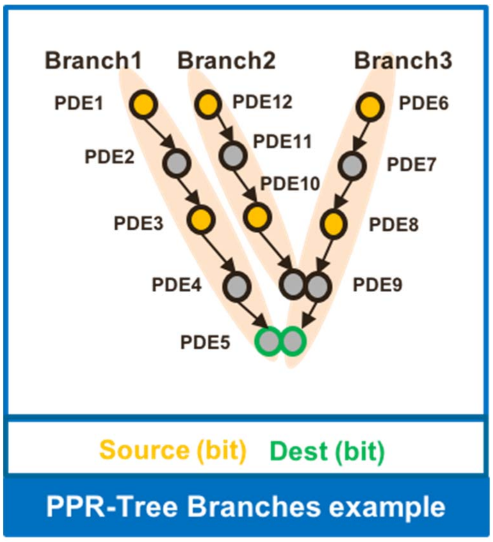

Figure 15: PPR Tree Graph representation

Rules:

- Every PDE-ID can be non-last-PDE in at most one Branch. It can be last-PDE in one or more Branches (e.g. PDE9).
- Branches form a tree by joining nodes with same PDE-ID (PDE9 and PDE5 in example). Leaves of the tree will be S(ources), e.g.: PDE1, PDE12, PDE6. Root of the tree will be the only D(estination) of the tree (e.g. PDE5).

How to build forwarding entry (referring to Figure 16 "PPR-Tree Branches Example" picture below):

- If PPR-ID in header of PPR is indicating this node (example: PDE5):
  - This node is D(estination) of this tree. Forwarding/Punt state is built for this PPR-Tree like for PPR-Path, no changes.
- If the PPR-ID is NOT indicating this node, then this node MAY be source (PDE1, PDE8) or midpoint (PDE9, neither source nor destination):
  - The processing node sequentially examines all branches until it finds a PDE with its own PDE-ID. It then establishes a forwarding entry for the PPR-ID indicated in the PPR header with the next-hop being the next PDE in the current branch.
  - This node's PDE may be the last PDE in a Branch, for example PDE9 in Branch9. In this case, the node ignores this branch because it cannot build a complete forwarding entry from it. Instead, it will build the forwarding entry from another branch, e.g.: Node with PDE9 will build forwarding entry for destination PDE5 when it examines Branch3 because there it will have a next hop PDE5. After forwarding entry is built, node can stop examining rest of Branch or further Branches.
  - If node does not find its own PDE in any branch it is not on the tree and ignores this PPR-Tree.
  - If node has S(ource) bit set in PDE, it can use this PPR to inject traffic.

Example 2 as shown below with 3 ingress PEs (Rs, R6 and R8) and 2 egress PEs (R13 and Rd). With a PPR Tree structure it gets simplified to one PPR as shown in the left with main path and branches as mentioned. In this case flooding optimization is achieved but all 3 paths need 3 separate FIB entries in the path nodes.

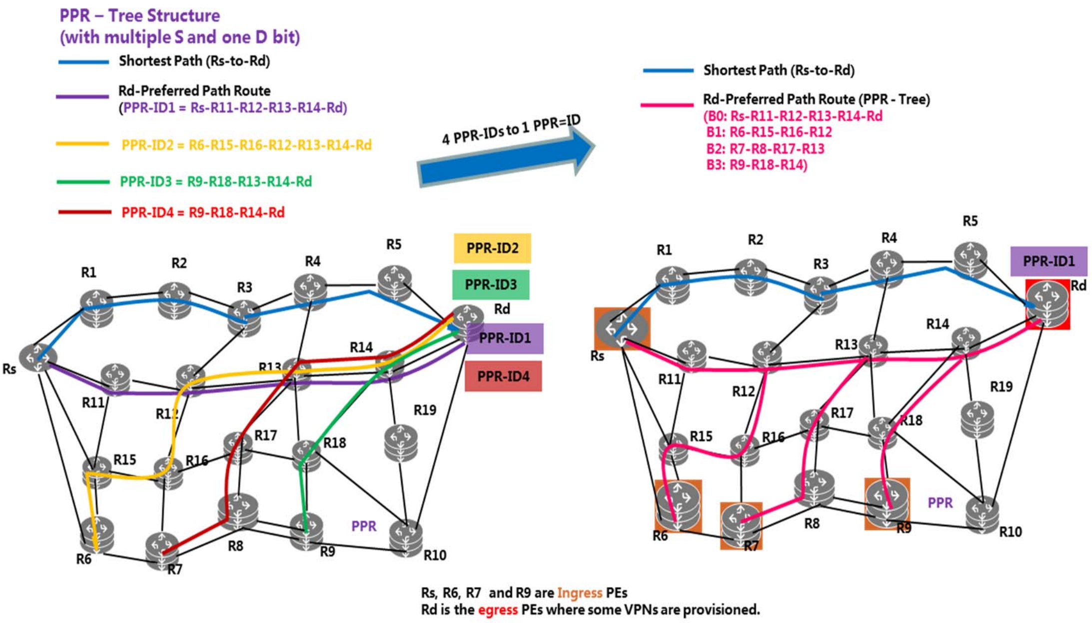

Figure 16: PPR Scalability with Tree Structure (example)

7.5 PPR Forests unidirectional and bidirectional

Forests are graphs with two or more destinations and one or more sources. Because forests can be seen as a more compressed form to represent multiple Trees, they are called Forests.

As shown in Figure 14 above, an unambiguous unidirectional and contiguously connected forest graph can have at most one loop. If it had two or more loops it would be ambiguous, or would consist of two or more disconnected forests which should then better be represented as separate forests.

A bidirectional forest is the only graph type in which the edges are meant to be bidirectional useable. Bidirectional Forests cannot have loops or else they would be ambiguous. Bidirectional forests can of course have more than one outgoing edge from a node, but those edges can only connect to disjoint set of destinations.

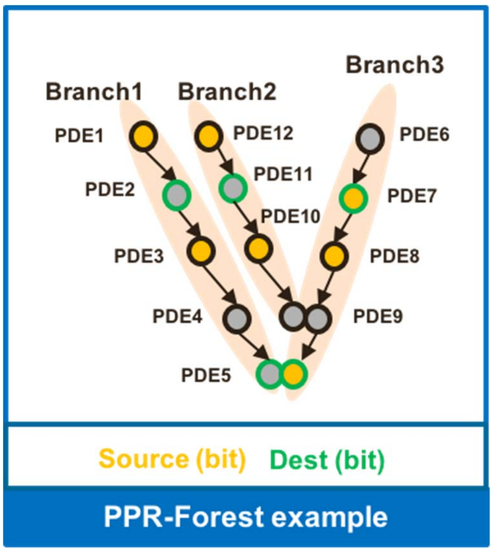

Figure 16a: PPR (unidirectional) Forest (example)

Parsing of PPR Forests is similar to parsing of PPR Trees and will not be explained in as much. The main added degree of work is that a node parsing a PPR Forest needs to establish a PPR-ID forwarding entry for every destination that is reachable from itself in the Forest. This is less work than parsing a separate Tree for each of those destinations. In the above example Figure 16a, PDE12 can reach destinations PDE11 and PDE5, but not PDE2 or PDE7. For every outgoing edge of the node itself, the node needs therefore to determine the set of reachable destinations and establish forwarding entries for those PPR-IDs across the adjacency represented by the edge.

Forests are a particularly elegant way to represent forwarding in rings because only two Forests are required to represent all possible paths between nodes in a ring, as shown in Figure 17.

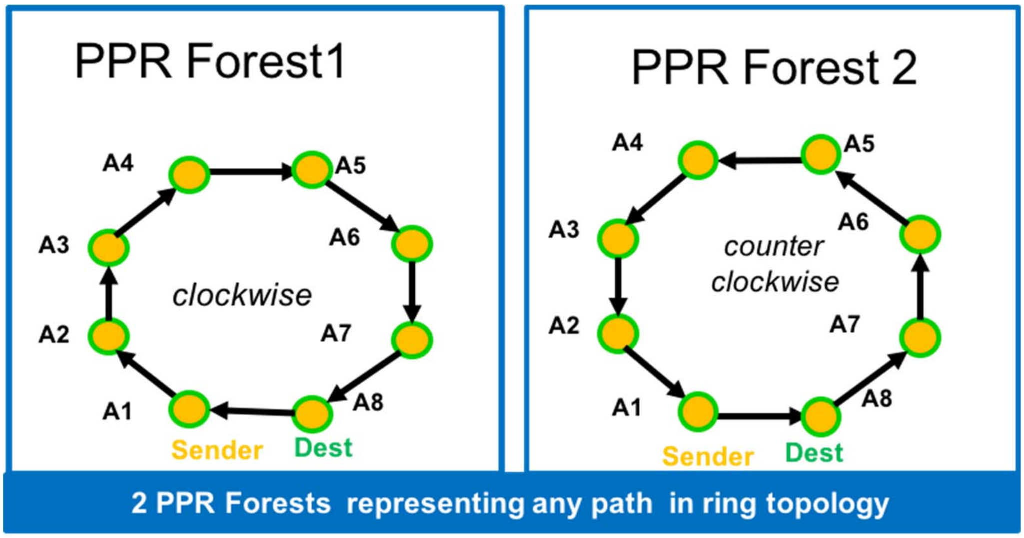

Figure 17: Usage of PPR Forest structure for RING Topologies

In this example, two forests represent a ring. All nodes have sender and destination bits set. A total of 16 PPR-ID are required: One for each node in Forest1 (clockwise), and one for each node in Forest2 (counter clockwise). This is equivalent to 2 × 8 × 7 = 112 PPR paths (and therefore PPR-IDs) - from any sender to any destination in either direction across the ring.

To encode a loop as happening in rings or other unidirectional PPR Forests with a loop, it would be possible to pick a random starting node, such as A1 and encode a branch starting from A1 going all the way to A8 but finishing up with A1 to indicate the loop. Instead of this duplicating the first to the last node in a branch, the PPR graph encoding simply supports an additional flag in the PPR branch data structure to indicate a looped branch.

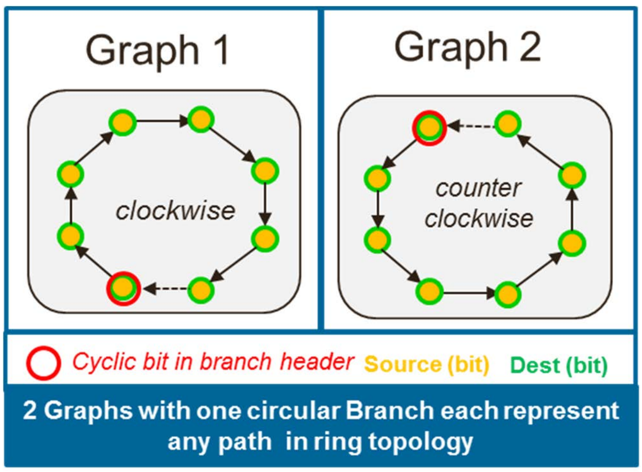

Figure 18: PPR with Cyclic Bit Example

A forest is not necessarily representing paths from all sources in the forest to all destinations in the forest, but only a subset that can be represented as a connected, non-ambiguous graph. As with the prior example of multiple trees, a complete set of paths between a set of sources and destinations may require multiple Trees and or Forests.

7.6 Creating Trees and Forests

Path computation Engines (PCEs) typically calculate just p2p paths. Once they have determined a set of paths, they can simply merge all feasible paths into PPR Trees, using multiple trees to a single destination as necessary. This step does not change the resulting traffic engineering paths but simply reduces the amount of PPR information that needs to be signalled, the number of forwarding entries (PPR-IDs) required, and the amount of (CPU) processing and memory required in the. These improvements all contribute to faster processing/convergence of PPR graphs and higher scalability.

If PPR graphs are just added as an optimization step without impacting the actual graph calculation, the achievable optimization may not be the best one. When the use case is for example global optimization of paths between all PE in a large service provider network, paths from different sources to the same destinations may converge earlier or later. The earlier these paths converge the better they can be merged into fewer trees or even Forests. When searching for the optimum set of paths, it may therefore be possible to take this additional optimization aspect into consideration and calculate path sets that can be better compressed into fewer trees or forests.

8 PPR Use Cases

8.1 Overview

The clause describes the various deployment use cases for PPR or PPR-TE with resource reservation along the path.

8.2 Access & Backhaul Transport Networks

In the current cellular architecture, traditional approaches like MPLS with RSVP-TE determine the Label Switched Path (LSP) and manage resources along the path. However, they are not dynamic and require provisioning steps along the path, involving out-of-band signalling. Another mechanism being considered is to use Segment Routing (SR) which is a source routing technology. SR (IETF RFC 8402 [i.17]), uses path segments computed offline by a controller for a particular flow or a service. The path is then decomposed into a sequence of network segments along which packets of a flow are routed. A sequence of segments is carried in the packet, essentially using source routing, with either MPLS labels or IPv6 address formats. While SR allows packet steering on a specified path, it does not have any notion of QoS or resources being reserved along the path. Furthermore, SR also has well-known overhead of increasing the packet header by encoding the segments into packets for routing purposes. This is explained further with 4G/LTE and 5G networks below and PPR here helps to achieve both lean packet over head as well as optional QoS attributes to the signalled path.

Figure 19 shows a typical 4G/LTE network with protocols stack ash shown at various points from UE to gNB, Serving and Packet Gateways. A transport backhaul network is defined from gNB to S-GW/P-GW in the core network.

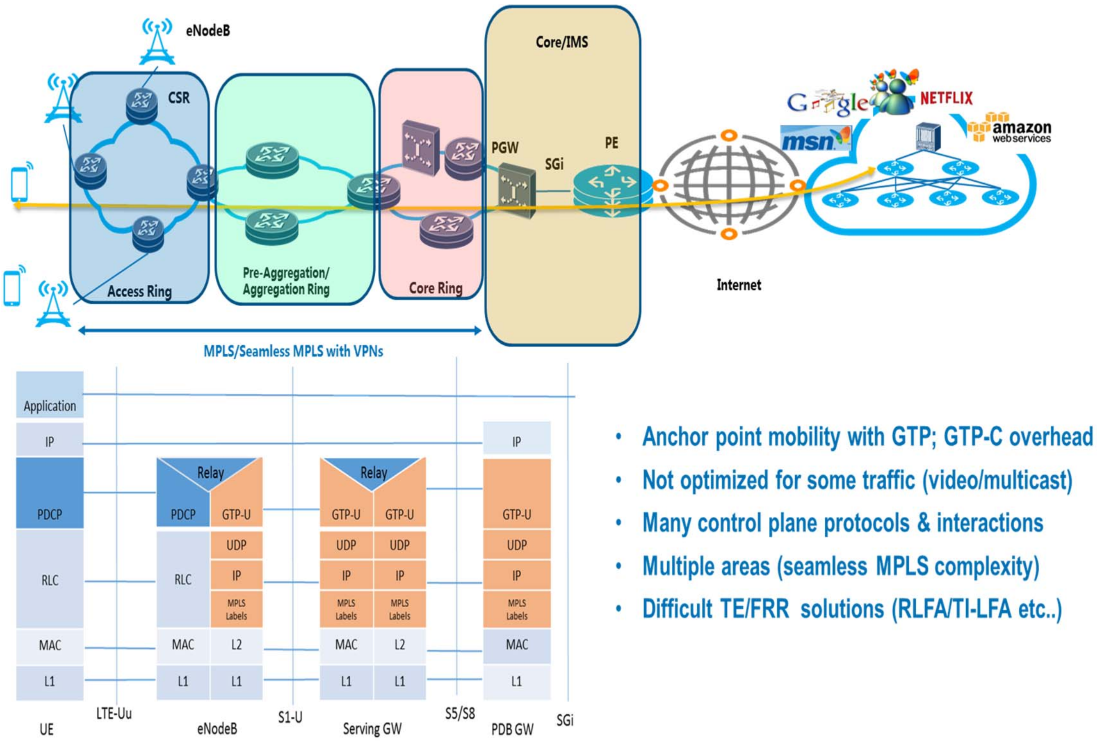

Figure 19: Typical 4G/LTE Transport Network with Protocol Stack

While there are some efforts to remove GTP in 5G, REL15 still continue to have anchor based GTP solution and this continues to be there on some interfaces in REL16. There are lot of enhancements have been done in the radio side for 5G with NR (New Radio) and fully virtualized packet core with decomposed core functions, transport network is not fully evolved for lot of 5G services, especially for uRLLC and mIOT. These services require matching service characteristics from the transport networks with optimized data plane both in steady state and during mobility, so as to not to erase the gains made in radio and core networks. PPR with optional QoS attributes fulfil the needs for these service while giving flexibility for the operators to use their choice of data planes. PPR provides both source routing capability with the centralized control plane as well an optimized data plane (regardless of the choice of data plane).

Figure 20 shows the evaluation of 5G where PPR have a role in the network. While a lot of proposals are being looked into by various forums on edge compute in 5G, it is not fully certain the edge would be move to gNB, mainly because of security concerns both physical, access and shared transport at gNB in some deployments. Though the pictures show the edge DC near gNB, it is possible can be eventually located at access/aggregation ring where some aggregation can happen and have better security perimeter. PPR QoS attributes are needed for the session terminating at the edge.

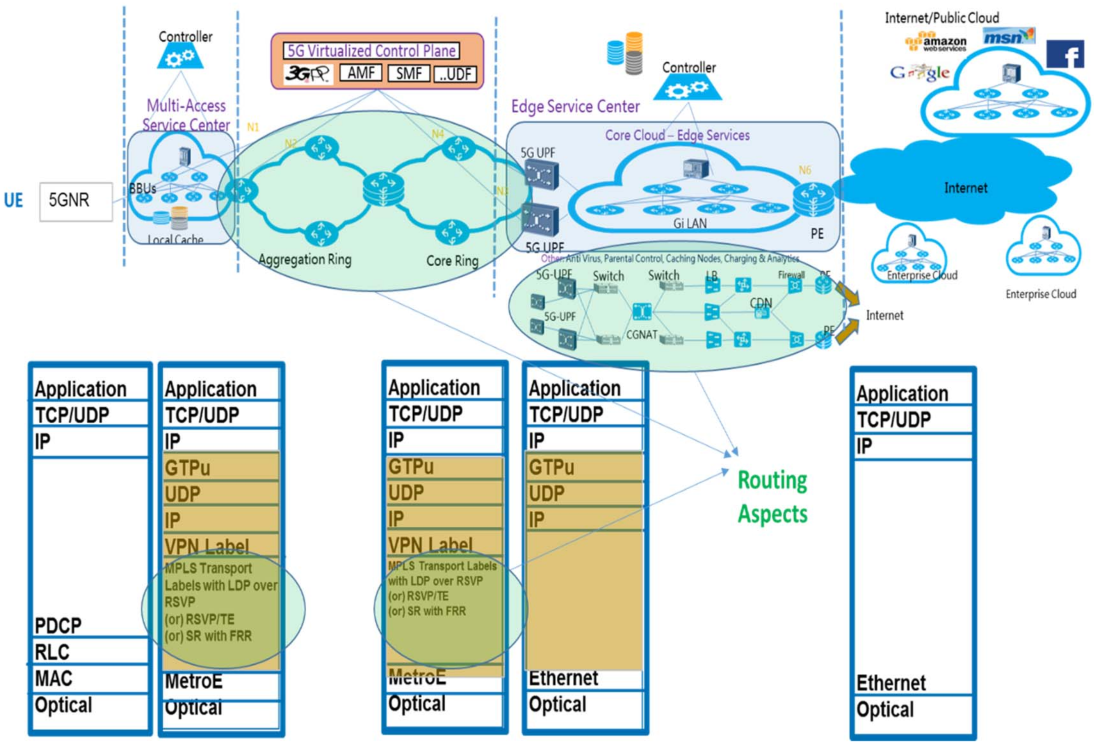

Figure 20: 5G Transport Network with Protocol Stack

PPR-IDs with different QoS characteristics for various slices, need to be setup for UPF nodes for uplink direction and gNB/Cell Site Routers (CSRs) for the downlink direction. Also a new control plane function needs to be present in 5G control plane to monitor the underlying transport network and creating the TE tunnels (SR or PPR) at both gNB/CSR and UPFs. Once this is done based on 5QI/QFI parameter in the uplink direction and RQI of the encapsulated packet in down link direction appropriate mapping of the underlying TE tunnel (PPR-ID) has to be done.

8.3 DC Underlays

With edge computing applications and the need for low latency service delivery micro and mini data centres are being deployed both in IP access networks and cellular font-haul and back-haul networks. Though at this point of time the industry is looking for the answer where the "edge" is, it could be close to gNB in 5G (physical security issues) or pre-aggregation or aggregation rings depending on the application and local caching needs for low latency services. Both the mini DCs and Massively Scalable Data Centres (MSDCs) switching to Layer3 based routing underlay managed by a central controller, the need for light weight path steering mechanism is an important tool for prioritizing the traffic as per the operator need.

The main areas where PPR can be used include Data Centre Interconnect (DCI), DC underlays and DC overlay networks or virtual network with virtual routers running on DC server farm. For the DCI scenario, PPR helps provide Traffic Engineering, optionally with resource reservations for high volume inter-DC traffic with no path overhead. PPR also provides needed flexibility for creating dynamic TE paths for the deployments, where native IP data planes (IPv4/IPv6) are used in most of the underlay deployments. With this operators can avoid the congestion in the underlay for critical servers dynamically. In the overlay or virtual router environment PPR provides light weight service chaining with non-topological PDEs along with the preferred path. PPR also help achieve OAM capabilities at the path granularity without any additional overhead in the traffic unlike iOAM technology.

8.4 SP and Access Networks

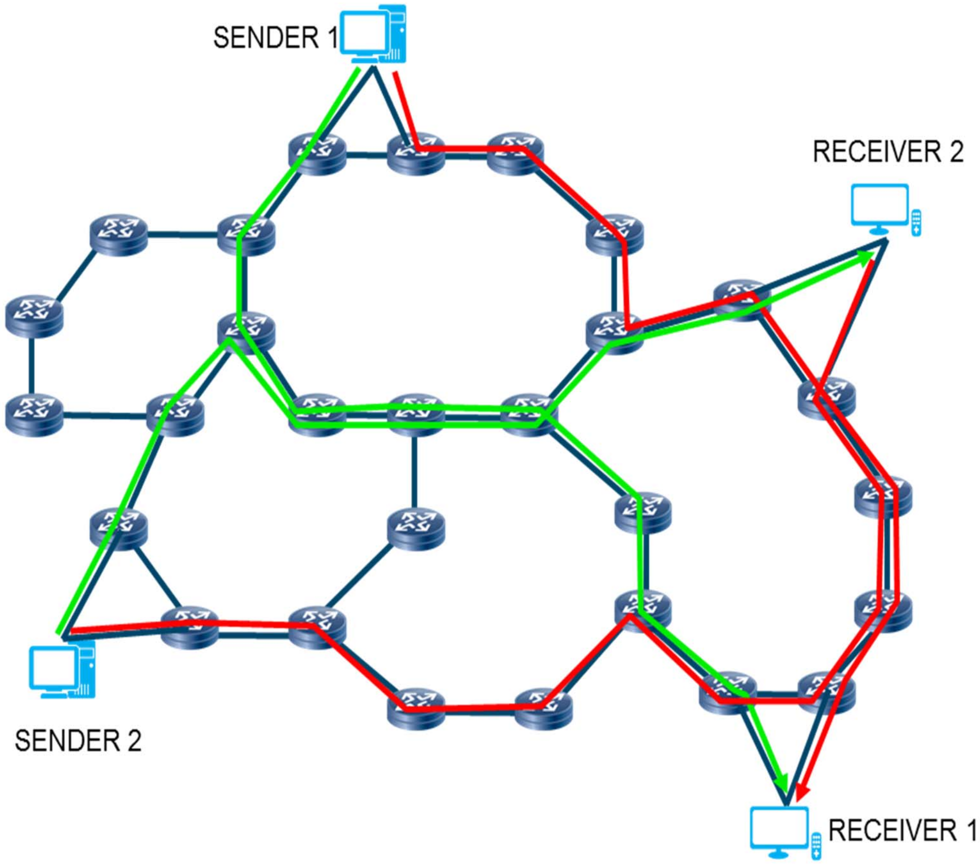

Figure 21: Sub-tended Ring topology example in access networks

Access networks are often characterized by the most cost optimized redundant network topologies in a wide area network such as so-called "market-networks" spanning metropolitan areas as large as 200 miles in diameter. The shortest total required connectivity is therefore often a complex topology of interconnected, so-called "sub-tended" rings as shown in above example. Services with a need of high reliability often use path redundancy, where the two alternative paths from a source to a destination do not share single points of failures. This is often preferred over node-local fast-reroute approaches because it allows not only the active switchover from one path to the other at the head end (upon failure of the first path), but also the so-called dual-path forwarding where both paths are actively used. This for example is required in high resilient services such as application that utilize Deterministic Networking services.

In simple rings this path redundancy is just clockwise vs. counter clockwise forwarding of traffic (as shown in Figure 21), but in real networks with subtended rings, this is a complex path engineering problem requiring the setup of strict hop-by-hop paths across paths with often as much as 20 or more hops. Compared to other traffic steering mechanisms, PPR is ideal for this type of use-cases because it is agnostic to the path length and it implicitly also supports a key functionality required in these setups: the non-rerouting of failed paths. In other technologies such as Segment Routing, it is difficult if not impossible to avoid re-routing of explicitly steered paths, which then often leads to that rerouted traffic to unnecessarily reduce the amount of resources available on the second set of paths.

This is further explained with the below example.

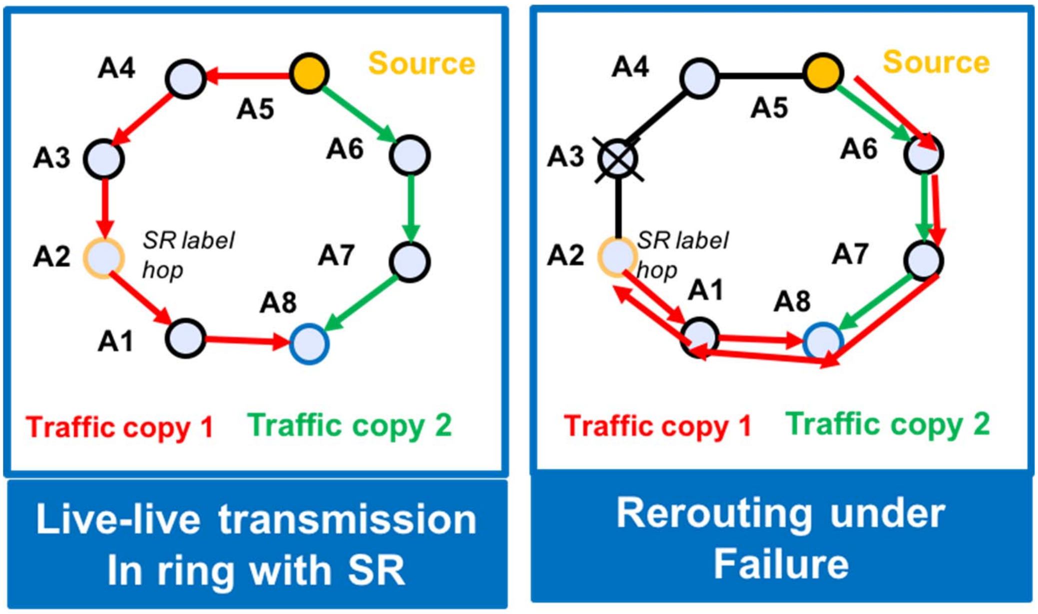

Figure 22: Example of dual-transmission in Ring topologies with Segment Routing

Consider dual-transmission in a ring with Segment Routing as shown in Figure 22 above on the left. To minimize the overhead of the SR SID stack in each packet it would be sufficient to insert no additional SID for the (green) traffic copy 2 from source to destination, because it would follow the shortest path routing. For (red) traffic copy 1, one additional SID would be required to steer packets away from the shortest path into the counter clockwise direction. Assume this hop is A2.

When as shown on the right side picture, A3 fails, the traffic copy 1 packets would follow the path as shown in the clockwise direction up to A2 because that is an explicitly indicated SID hop. And from there back to the destination A8.This rerouted traffic would not only consume in an often undesired fashion capacity in the clockwise direction from A5 to A8, but also in both directions between A2 and A8.

To avoid this overhead and rerouting, it would be necessary in SR to include full hop-by-hop SID stacks for both traffic copy packets and also ensure that those SID adjacencies are not subject to rerouting (via appropriate configuration). With rings often having a large number of hops, this is infeasible in many real deployments. Compared to this as explained, PPR is ideal for this type of use-cases because it is agnostic to the path length and it implicitly also supports a key functionality required in these setups.

9 Summary

Preferred Path Routing as described in the present document claims several benefits over existing systems (Segment Routing or RSVP-TE) for 5G and beyond. Furthermore, PPR infrastructure is an enabler for a host of new services as well as network management. Lastly, the PPR solutions can seamlessly interoperate with the current architecture and yet be flexible enough to interoperate with future forwarding infrastructures as well.

Annex A: Bibliography

ETSI GS NGP 001: "Next Generation Protocols (NGP); Scenario Definitions".

ETSI TS 129 060: "Digital cellular telecommunications system (Phase 2+) (GSM); Universal Mobile Telecommunications System (UMTS); General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface (3GPP TS 29.060)".

IETF RFC 7752: "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, March 2016.

IEEE Internet Computing: "Improving Performance and Scalability of Next Generation Cellular Networks", A. Mohammadkhan, K. K. Ramakrishnan, U. Chunduri, K. Makhijani, 2019 Volume: 23 Issue: 1.

Annex B:
Authors & contributors

The following people have contributed to the present document:

Rapporteur:

Uma Chunduri, Futurewei Technologies uma.chunduri@futurewei.com

Contributors:

Toerless Eckert, Futurewei Technologies toerless.eckert@futurewei.com

Yingzhen Qu, Futurewei Technologies yingzhen.qu@futurewei.com

Alexander Clemm, Futurewei Technologies alclemm@futurewei.com

Richard Li, Futurewei Technologies richard.li@futurewei.com

Annex C: Change History

Date	Version	Information about changes
June 7th, 2017	0.0.0	WI#14 approved and document created
Aug 31st 2018	0.0.1	Content added to PPR paths, graphs and use cases
Nov 30th 2018	0.0.2	Added other PPR graph types and various use cases with PPR
Apr 17th 2019	0.0.3	Additional reference and updated MSDC use case
June 6th 2019	1.0.0	ETSI Compliance Review
Oct 4th 2019	1.1.1	ETSI Final Review comments addressed

History

Document history		
V1.1.1	October 2019	Publication