

MPLS Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: September 11, 2017

H. Sitaraman  
V. Beeram  
Juniper Networks  
T. Parikh  
Verizon  
March 10, 2017

Signaling RSVP-TE tunnels on a shared MPLS forwarding plane  
draft-sitaraman-mpls-rsvp-shared-labels-00.txt

Abstract

As the scale of MPLS RSVP-TE LSPs has grown, various implementation recommendations have been proposed to manage control plane state. However, the forwarding plane footprint of labels at a transit LSR has remained proportional to the total LSP state in the control plane. This draft defines a mechanism to prevent the label space limit on an LSR from being a constraint to control plane scaling on that node. It introduces the notion of pre-installed per TE link 'pop labels' that are shared by MPLS RSVP-TE LSPs that traverse these links and thus significantly reducing the forwarding plane state required. This couples the feature benefits of the RSVP-TE control plane with the simplicity of the Segment Routing MPLS forwarding plane. This document also introduces the ability to mix different types of label operations along the path of the LSP, thereby allowing the ingress or an external controller to influence how to optimally place a LSP.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on September 11, 2017.

## Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

## Table of Contents

1. Introduction . . . . .	3
2. Conventions used in this document . . . . .	4
3. Terminology . . . . .	4
4. Allocation of pop labels . . . . .	4
5. RSVP-TE pop and forward tunnel setup . . . . .	4
6. Mixing pop and swap labels in a RSVP-TE tunnel . . . . .	6
7. Distributing label stack imposition . . . . .	7
8. Facility backup protection . . . . .	7
8.1. Link Protection . . . . .	7
8.2. Node Protection . . . . .	8
9. Quantifying pop labels . . . . .	8
10. Protocol Extensions . . . . .	9
10.1. Requirements . . . . .	9
10.2. Attributes Flags TLV: Pop Label . . . . .	9
10.3. RRO Label Subobject Flag: Pop Label . . . . .	10
10.4. Attributes TLV: Label Stack Imposition TLV . . . . .	10
11. OAM considerations . . . . .	11
12. Acknowledgements . . . . .	11
13. Contributors . . . . .	11
14. IANA Considerations . . . . .	11
14.1. Attribute Flags: Pop Label . . . . .	11
14.2. Attribute TLV: Label Stack Imposition TLV . . . . .	11
14.3. Record Route Label Sub-object Flags: Pop Label . . . . .	12
15. Security Considerations . . . . .	12
16. References . . . . .	12
16.1. Normative References . . . . .	12
16.2. Informative References . . . . .	13
Authors' Addresses . . . . .	14

## 1. Introduction

Various RSVP-TE scaling recommendations [RFC2961] [I-D.ietf-teas-rsvp-te-scaling-rec] have been proposed for implementations to adopt guidelines that would allow the RSVP-TE [RFC3209] control plane to scale better. The forwarding plane state required to handle the equivalent control plane state remains unchanged and is proportional to the total LSP state in the control plane. The motivation of this draft is to prevent the platform specific label space limit on an LSR from being a constraint to pushing the limits of control plane scaling on that node.

This document proposes the allocation of a 'pop label' by a LSR for each of its TE links. The label is installed in the MPLS forwarding plane with a pop label operation and to forward the received packet over the TE link. This label is sent normally by the LSR in the Label object in the Resv message as LSPs are setup. The ingress LER SHOULD construct and push a stack of labels [RFC3031] as received in the Record Route object (RRO) in the Resv message.

This pop and forward data plane behavior is similar to that used by Segment Routing (SR) [I-D.ietf-spring-segment-routing] using a MPLS forwarding plane and a series of adjacency segments. The RSVP-TE pop and forward tunnels can co-exist with SR LSPs as described in [I-D.sitaraman-sr-rsvp-coexistence-rec].

RSVP-TE using a pop and forward data plane offers the following benefits:

1. Shared forwarding plane: The transit label on a TE link is shared among RSVP-TE tunnels traversing the link and is used independent of the ingress and egress of the LSPs.
2. Faster LSP setup time: The forwarding plane state is not programmed during LSP setup and teardown resulting in faster LSP setup time.
3. Hitless routes: New transit labels are not required on complete path overlap during make-before-break (MBB) resulting in a faster MBB event. This avoids the ingress LER and the services that might be using the tunnel from needing to update its forwarding plane with new tunnel labels. Periodic MBB events are relatively common in networks that deploy auto-bandwidth on RSVP-TE LSPs to monitor bandwidth utilization and periodically adjust LSP bandwidth.
4. Mix and match labels: Both 'pop' and 'swap' labels can be mixed across transit hops for a single RSVP-TE tunnel (see Section 6).

This allows local policy at an ingress or path computation engine to influence RSVP-TE to mix and match different types of labels across a LSP path.

No additional extensions are required to IGP-TE in order to support this pop and forward data plane. Functionalities such as bandwidth admission control, LSP priorities, preemption, auto-bandwidth and Fast Reroute continue to work with this forwarding plane.

## 2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

## 3. Terminology

Pop label: An incoming label at a LSR that will be popped and forwarded over a specific TE link to a neighbor.

Swap label: An incoming label at a LSR that will be swapped to an outgoing label and forwarded over a specific downstream TE link.

Pop and forward data plane: A forwarding plane where every LSR along the path uses a pop label.

RSVP-TE pop and forward tunnel: A MPLS RSVP-TE tunnel that uses a pop and forward data plane.

## 4. Allocation of pop labels

A LSR SHOULD allocate a unique pop label for each TE link. The forwarding action for the pop label should it appear on top of the label stack MUST be to pop the label and forward the packet over the TE link to the downstream neighbor of the RSVP-TE tunnel. Multiple labels MAY be allocated for the TE link to accommodate tunnels requesting no protection, link-protection and node-protection over the specific TE link.

## 5. RSVP-TE pop and forward tunnel setup

This section provides an example of how the RSVP-TE signaling procedure works to setup a tunnel utilizing a pop and forward data plane. The sample topology below will be used to explain the setup.

Labels shown at each node are pop labels for that neighbor

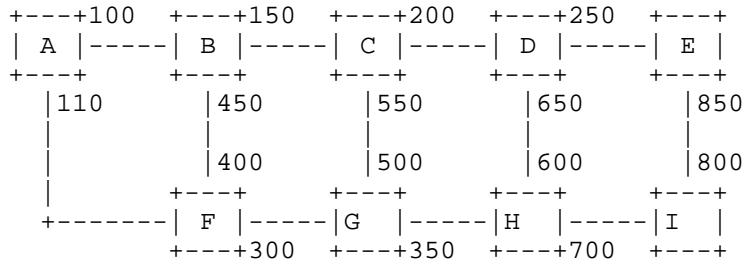


Figure 1: Pop and forward label topology

RSVP-TE tunnel T1: From A to E on path A-B-C-D-E

RSVP-TE tunnel T2: From F to E on path F-B-C-D-E

Both tunnels share the TE links B-C, C-D and D-E.

As RSVP-TE signals the setup (using the pop label attributes flag defined in Section 10.2) of tunnel T1, when LSR D receives the Resv message from the egress E, it checks the next-hop TE link (D-E) and provides the pop label (250) in the Resv message for the tunnel. The label is sent in the Label object and is also recorded in the Label sub-object (using the pop label bit defined in Section 10.3) carried in the RRO. Similarly, C provides the pop label (200) for the next-hop TE link C-D and B provides the pop label (150) for the next-hop TE link B-C. For the tunnel T2, the transit LSRs provide the same pop labels as described for tunnel T1.

Both LER A and F will push the same stack of labels {150(top), 200, 250} for tunnels T1 and T2 respectively. It should be noted that a transit LSR does not use the pop label provided in the label object by its downstream LSR in the NHLFE as the outgoing label. The recorded labels in the RRO are of interest to the ingress LER in order to construct a stack of labels.

If there were another RSVP-TE tunnel T3 from F to I on path F-B-C-D-E-I, then this would also share the TE links B-C, C-D and D-E and additionally traverse link E-I. The label stack used by F would be {150(top), 200, 250, 850}. Hence, regardless of the ingress and egress LERs from where the LSPs start and end, they will share LSR labels at shared hops in the pop and forward data plane.

There MAY be local operator policy at the ingress LER that influences the maximum depth of the label stack that can be pushed for a RSVP-TE pop and forward tunnel. Prior to signaling the LSP, if the ingress LER decides that it would be unable to push the entire label stack

should every transit hop provide a pop label, then the LER can choose to either not signal a RSVP-TE pop and forward tunnel or can adopt techniques mentioned in Section 6 or Section 7.

## 6. Mixing pop and swap labels in a RSVP-TE tunnel

Labels can be mixed across transit hops in a single MPLS RSVP-TE LSP. Certain LSRs can use pop labels and others can use swap labels. The ingress can construct a label stack appropriately based on what type of label is recorded from every transit LSR.

Labels shown at each node are pop labels for that TE link. (#) are swap labels.

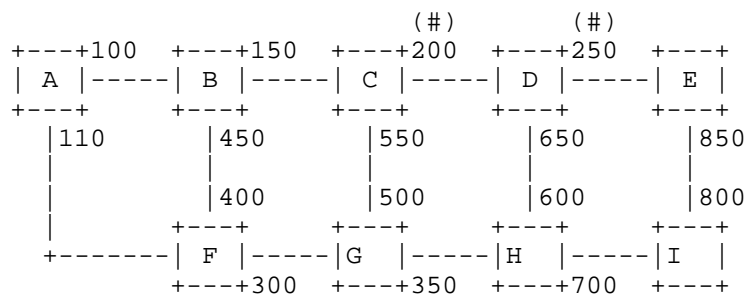


Figure 2: Mix pop and swap label topology

If the transit LSR is allocating a swap label to be sent upstream in the Resv, then the label operation in the NHLFE MUST be a swap to any label received from the downstream LSR. If the transit LSR is using a pop label to be sent upstream in the Resv, then the label operation in the NHLFE MUST be a pop and forward regardless of any label received from the downstream LSR.

The ingress LER MUST check the type of label received from each transit hop as recorded in the RRO in the Resv message and generate the appropriate label stack to use for the RSVP-TE tunnel.

The following logic could be used by the ingress LER while constructing the label stack:

Each RRO label sub-object SHOULD be processed starting with the label sub-object from the first downstream hop. Any label provided by the first downstream hop MUST always be pushed on the label stack regardless of the label type. If the label type is a pop label, then any label from the next downstream hop MUST also be pushed on the constructed label stack. If the label type is a swap label, then any label from the next downstream hop MUST NOT be pushed on the

constructed label stack. For example, the LSP from A to I using path A-B-C-D-E-I will use a label stack of {150(top), 200}.

Signaling extensions for the ingress LER to request a certain type of label from a particular hop is defined in Section 10.2. A Hop-Count value of 1 (Label Stack Imposition Attribute) SHOULD be used for the specific hops to allocate a swap label.

## 7. Distributing label stack imposition

One or more transit LSRs can assist the ingress LER by imposing part of the label stack required for the path. From Figure 1, ingress LER A can use the assistance of transit LSRs to push labels downstream of that LSR. For example, LER A can push label 150 and LSR C can push {200(top), 250} for the LSP taking path A-B-C-D-E.

The ingress LER can request one or more specific transit hops to handle pushing labels for N of its downstream hops. To achieve this request properly, the ingress can learn the label stack depth push limit of the transit LSRs. The mechanism by which the ingress or controller (hosting the path computation element) learns this information is outside the scope of this document. The particular transit hops SHOULD allocate a swap label that will result in that label being replaced and a set of labels pushed to accommodate N downstream hops.

Signaling extensions for the ingress LER to request one or more transit LSRs to handle label stack imposition for N downstream hops or for the transit hop to indicate to the ingress that it can handle label stack imposition for N downstream hops is defined in Section 10.2. The Hop-Count field (Label Stack Imposition Attribute) can be used to indicate the value of N.

## 8. Facility backup protection

The following section describe how link and node protection works with facility backup protection [RFC4090] for the RSVP-TE pop and forward tunnels.

### 8.1. Link Protection

To provide link protection at a PLR with a pop and forward data plane, the LSR SHOULD allocate a separate pop label for the TE link that will be used for RSVP-TE tunnels that request link-protection from the ingress. No signaling extensions are required to support link protection for RSVP-TE tunnels over the pop and forward data plane.

(\*) are pop labels to offer link protection for that TE link

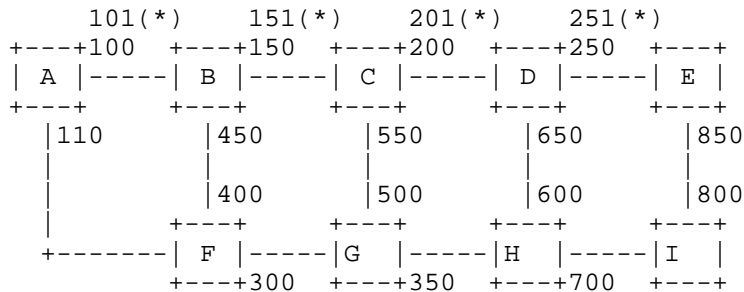


Figure 3: Link protection topology

At each LSR, link protected pop labels can be allocated for each TE link and a link protecting facility backup LSP can be created to protect the TE link. This label can be sent by the LSR for LSPs requesting link-protection over the specific TE link. Since the facility backup terminates at the next-hop (merge point), the incoming label on the packet will be what the merge point expects.

As an example, LSR B can install a facility backup LSP for the link protected pop label 151. When the TE link B-C is up, LSR B will pop 151 and send the packet to C. If the TE link B-C is down, the LSR can pop 151 and send the packet via the facility backup to C.

## 8.2. Node Protection

The solutions for the PLR to provide node-protection for the pop and forward RSVP-TE tunnel will be explained in the next version of the document.

## 9. Quantifying pop labels

This section attempts to quantify the number of labels required in the forwarding plane to provide sharing of labels across RSVP-TE pop and forward tunnels. A MPLS RSVP-TE tunnel offers either no protection, link protection or node protection and only one of these labels is required per tunnel during signaling. The scale of the number of pop labels required per LSR can be deduced as follows:

- o For a LSR having X neighbors reachable across Y interfaces, the number of unprotected pop labels = X



- o For a PLR having X neighbors reachable across Y interfaces, number of link protected pop labels = X
- o For a PLR having X neighbors, each having Nx neighbors (i.e. next-nexthop for PLR), number of node protected pop labels = SUM\_OF\_ALL(Nx)

Total number of pop labels = Unprotected pop labels + link protected pop labels + node protected pop labels = 2X + SUM(Nm)

## 10. Protocol Extensions

### 10.1. Requirements

The functionality discussed in this document imposes the following requirements on the signaling protocol.

- o The Ingress of the LSP SHOULD have the ability to mandate/request the use and recording of pop labels at all hops along the path of the LSP.
- o When the use of pop labels is mandated/requested for the entire path,

the node recording the pop label SHOULD have the ability to indicate if the recorded label is a pop label.

the ingress SHOULD have the ability to override this path specific behavior by

explicitly mandating specific hops to not use pop labels (or)

mandating specific hops to share the onus of imposing the label stack (and also specifying the desired number of hops that need to be accounted for at that node)

the node which was mandated to share the onus of imposing the label stack SHOULD have the ability to indicate the actual number of hops that it can account for.

### 10.2. Attributes Flags TLV: Pop Label

Bit Number (TBD1): Pop Label

The presence of this in the LSP\_ATTRIBUTES/LSP\_REQUIRED\_ATTRIBUTES object of a Path message indicates that the ingress has requested/mandated the use and recording of pop labels at all hops along the

path of this LSP. When a node that does not cater to the request/mandate receives a Path message carrying the LSP\_REQUIRED\_ATTRIBUTES object with this flag set, it MUST send a PathErr message with an error code of 'routing problem' and an error value of 'pop label usage failure'.

### 10.3. RRO Label Subobject Flag: Pop Label

Bit Number (TBD2): Pop Label

The presence of this flag indicates that the recorded label is a pop label. This flag SHOULD be used by a node only if the use and recording of pop labels is requested/mandated for this LSP.

### 10.4. Attributes TLV: Label Stack Imposition TLV

0										1										2										3									
0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9
Reserved																				Hop-Count																			

Attribute TLV Type: TBD3

The presence of this in the HOP\_ATTRIBUTES subobject [RFC7570] of an ERO object in the Path message mandates the hop identified by the preceding IPv4 or IPv6 or Unnumbered Interface ID subobject to share the onus of imposing the label stack. This attribute MUST be used only if the use and recording of pop labels is requested/mandated for this LSP (only if Pop Label flag is present in the LSP\_ATTRIBUTES/LSP\_REQUIRED\_ATTRIBUTES object). If the node is not able to comply with this mandate, it MUST send a PathErr message with an error code of 'routing problem' and an error value of 'label stack imposition failure'.

The Hop-Count field specifies the desired number of hops that this node needs to account for. A Hop-Count value of 0 is considered invalid and a value of 1 implies that this hop perform a normal swap or pop (if this hop is PHP) operation towards the next downstream hop.

The presence of this in the HOP\_ATTRIBUTES subobject of an RRO object in the RESV message indicates that the hop identified by the preceding IPv4 or IPv6 or Unnumbered Interface ID subobject is sharing the onus of imposing the label stack. The Hop-Count field specifies the actual number of hops that this node can account for. This should not be included in the RESV message unless this TLV is also present in the corresponding Path message for this hop.

## 11. OAM considerations

Any extensions necessary for MPLS LSP traceroute for the RSVP-TE pop and forward tunnel will be explained in the next version of the document.

## 12. Acknowledgements

The authors would like to thank Adrian Farrel, Kireeti Kompella, Markus Jork and Ross Callon for their input from discussions.

## 13. Contributors

The following individuals contributed to this document:

Raveendra Torvi  
Juniper Networks  
Email: rtorvi@juniper.net

Chandra Ramachandran  
Juniper Networks  
Email: csekar@juniper.net

## 14. IANA Considerations

### 14.1. Attribute Flags: Pop Label

IANA manages the 'Attribute Flags' registry as part of the 'Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters' registry located at <http://www.iana.org/assignments/rsvp-te-parameters>. This document introduces a new Attribute Flag.

Bit No.	Name	Attribute Flags	Attribute Path	Attribute Resv	RRO	ERO	Reference
TBD1	Pop Label	Yes	No	No	No	No	This document (Section 5)

### 14.2. Attribute TLV: Label Stack Imposition TLV

IANA manages the "Attribute TLV Space" registry as part of the 'Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters' registry located at <http://www.iana.org/assignments/rsvp-te-parameters>. This document introduces a new Attribute TLV.

Type	Name	Allowed on LSP ATTRIBUTES	Allowed on LSP REQUIRED ATTRIBUTES	Allowed on LSP Hop Attributes	Reference
TBD3	Label Stack Imposition TLV	No	No	Yes	This document (Section 7)

#### 14.3. Record Route Label Sub-object Flags: Pop Label

IANA manages the 'Record Route Object Sub-object Flags' registry as part of the 'Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Parameters' registry located at <http://www.iana.org/assignments/rsvp-te-parameters>. This registry currently does not include Label Sub-object Flags. This document proposes the addition of a new sub-registry for Label Sub-object Flags as shown below.

Flag	Name	Reference
0x1	Global Label	RFC 3209
TBD2	Pop Label	This document (Section 5)

#### 15. Security Considerations

This document does not introduce new security issues. The security considerations pertaining to the original RSVP protocol [RFC2205] and RSVP-TE [RFC3209] and those that are described in [RFC5920] remain relevant.

#### 16. References

##### 16.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC2205] Braden, R., Ed., Zhang, L., Berson, S., Herzog, S., and S. Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1 Functional Specification", RFC 2205, DOI 10.17487/RFC2205, September 1997, <<http://www.rfc-editor.org/info/rfc2205>>.

- [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, DOI 10.17487/RFC3031, January 2001, <<http://www.rfc-editor.org/info/rfc3031>>.
- [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V., and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001, <<http://www.rfc-editor.org/info/rfc3209>>.
- [RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090, DOI 10.17487/RFC4090, May 2005, <<http://www.rfc-editor.org/info/rfc4090>>.
- [RFC7570] Margaria, C., Ed., Martinelli, G., Balls, S., and B. Wright, "Label Switched Path (LSP) Attribute in the Explicit Route Object (ERO)", RFC 7570, DOI 10.17487/RFC7570, July 2015, <<http://www.rfc-editor.org/info/rfc7570>>.

## 16.2. Informative References

- [I-D.ietf-spring-segment-routing]  
Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing-11 (work in progress), February 2017.
- [I-D.ietf-teas-rsvp-te-scaling-rec]  
Beeram, V., Minei, I., Shakir, R., Pacella, D., and T. Saad, "Implementation Recommendations to Improve the Scalability of RSVP-TE Deployments", draft-ietf-teas-rsvp-te-scaling-rec-03 (work in progress), October 2016.
- [I-D.sitaraman-sr-rsvp-coexistence-rec]  
Sitaraman, H., Beeram, V., Minei, I., and S. Sivabalan, "Recommendations for RSVP-TE and Segment Routing LSP co-existence", draft-sitaraman-sr-rsvp-coexistence-rec-02 (work in progress), February 2017.
- [RFC2961] Berger, L., Gan, D., Swallow, G., Pan, P., Tommasi, F., and S. Molendini, "RSVP Refresh Overhead Reduction Extensions", RFC 2961, DOI 10.17487/RFC2961, April 2001, <<http://www.rfc-editor.org/info/rfc2961>>.

[RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010, <<http://www.rfc-editor.org/info/rfc5920>>.

Authors' Addresses

Harish Sitaraman  
Juniper Networks  
1133 Innovation Way  
Sunnyvale, CA 94089  
US

Email: [hsitaraman@juniper.net](mailto:hsitaraman@juniper.net)

Vishnu Pavan Beeram  
Juniper Networks  
10 Technology Park Drive  
Westford, MA 01886  
US

Email: [vbeeram@juniper.net](mailto:vbeeram@juniper.net)

Tejal Parikh  
Verizon  
400 International Parkway  
Richardson, TX 75081  
US

Email: [tejal.parikh@verizon.com](mailto:tejal.parikh@verizon.com)