

Internet Engineering Task Force (IETF)  
Request for Comments: 9722  
Updates: 8584  
Category: Standards Track  
ISSN: 2070-1721

P. Brissette  
A. Sajassi  
LA. Burdet, Ed.  
Cisco  
J. Drake  
Independent  
J. Rabadan  
Nokia  
May 2025

## Fast Recovery for EVPN Designated Forwarder Election

### Abstract

The Ethernet Virtual Private Network (EVPN) solution in RFC 7432 provides Designated Forwarder (DF) election procedures for multihomed Ethernet Segments. These procedures have been enhanced further by applying the Highest Random Weight (HRW) algorithm for DF election to avoid unnecessary DF status changes upon a failure. This document improves these procedures by providing a fast DF election upon recovery of the failed link or node associated with the multihomed Ethernet Segment. This document updates RFC 8584 by optionally introducing delays between some of the events therein.

The solution is independent of the number of EVPN Instances (EVIs) associated with that Ethernet Segment, and it is performed via a simple signaling in BGP between the recovered node and each of the other nodes in the multihoming group.

### Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <https://www.rfc-editor.org/info/rfc9722>.

### Copyright Notice

Copyright (c) 2025 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 (<https://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 Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

### Table of Contents

#### 1. Introduction

- 1.1. Requirements Language
- 1.2. Terminology
- 1.3. Challenges with Existing Mechanism
- 1.4. Design Principles for a Solution
- 2. DF Election Synchronization Solution
  - 2.1. BGP Encoding
  - 2.2. Timestamp Verification
  - 2.3. Updates to RFC 8584
- 3. Synchronization Scenarios
  - 3.1. Concurrent Recoveries
- 4. Backwards Compatibility
- 5. Security Considerations
- 6. IANA Considerations
- 7. References
  - 7.1. Normative References
  - 7.2. Informative References

Acknowledgements

Contributors

Authors' Addresses

## 1. Introduction

The Ethernet Virtual Private Network (EVPN) solution [RFC7432] is widely used in data center (DC) applications for Network Virtualization Overlay (NVO) and Data Center Interconnect (DCI) services and in service provider (SP) applications for next-generation virtual private LAN services.

[RFC7432] describes Designated Forwarder (DF) election procedures for multihomed Ethernet Segments. These procedures are enhanced further in [RFC8584] by applying the Highest Random Weight (HRW) algorithm for DF election in order to avoid unnecessary DF status changes upon a link or node failure associated with the multihomed Ethernet Segment.

This document makes further improvements to the DF election procedures in [RFC8584] by providing an option for a fast DF election upon recovery of the failed link or node associated with the multihomed Ethernet Segment. This DF election is achieved independent of the number of EVPN Instances (EVIs) associated with that Ethernet Segment, and it is performed via straightforward signaling in BGP between the recovered node and each of the other nodes in the multihomed Ethernet Segment redundancy group.

This document updates the DF Election Finite State Machine (FSM) described in Section 2.1 of [RFC8584] by optionally introducing delays between some events, as further detailed in Section 2.3. The solution is based on a simple one-way signaling mechanism.

### 1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

### 1.2. Terminology

PE: Provider Edge

DF: Designated Forwarder. A PE that is currently forwarding (encapsulating/decapsulating) traffic for a given VLAN in and out of a site.

NDF: Non-Designated Forwarder. A PE that is currently blocking

traffic (see DF above).

EVI: EVPN Instance. It spans the PE devices participating in that EVPN.

HRW: Highest Random Weight algorithm [HRW98]

Service carving: This refers to DF election, as defined in [RFC7432].

SCT: Service Carving Time. Defined in this document as the time at which all nodes participating in an Ethernet Segment perform DF Election.

### 1.3. Challenges with Existing Mechanism

In EVPN technology, multiple PE devices encapsulate and decapsulate data belonging to the same VLAN. Under certain conditions, this may cause duplicated Ethernet packets and potential loops if there is a momentary overlap in forwarding roles between two or more PE devices, potentially also leading to broadcast storms of frames forwarded back into the VLAN.

EVPN [RFC7432] currently specifies timer-based synchronization among PE devices within an Ethernet Segment redundancy group. This approach can lead to duplications and potential loops due to multiple DFs if the timer interval is too short or can lead to packet drops if the timer interval is too long.

Split-horizon filtering, as described in Section 8.3 of [RFC7432], can prevent loops but does not address duplicates. However, if there are overlapping DFs of two different sites simultaneously for the same VLAN, the site identifier will differ when the packet re-enters the Ethernet Segment. Consequently, the split-horizon check will fail, resulting in Layer 2 loops.

The updated DF procedures outlined in [RFC8584] use the well-known HRW algorithm to prevent the reshuffling of VLANs among PE devices within the Ethernet Segment redundancy group during failure or recovery events. This approach minimizes the impact on VLANs not assigned to the failed or recovered ports and eliminates the occurrence of loops or duplicates during such events.

However, upon PE insertion or a port being newly added to a multihomed Ethernet Segment, the HRW cannot help either, as a transfer of the DF role to the new port must occur while the old DF is still active.

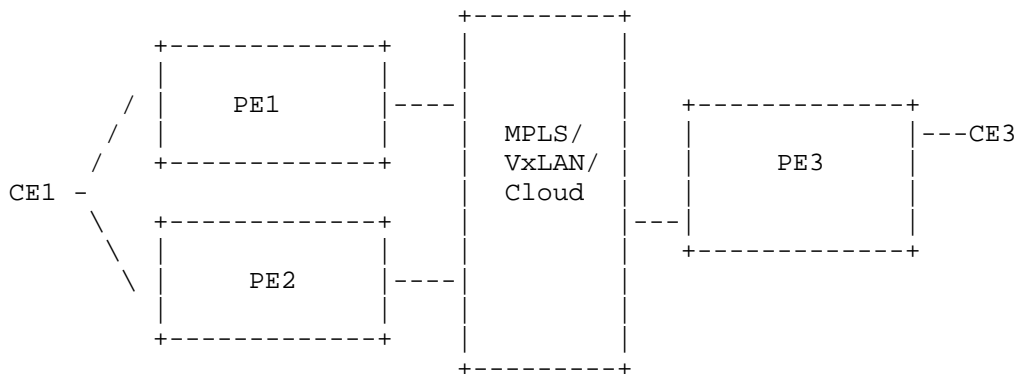


Figure 1: CE1 Multihomed to PE1 and PE2

In Figure 1, when PE2 is inserted in the Ethernet Segment or its CE1-facing interface is recovered, PE1 will transfer the DF role of

some VLANs to PE2 to achieve load-balancing. However, because there is no handshake mechanism between PE1 and PE2, overlapping of DF roles for a given VLAN is possible, which leads to duplication of traffic as well as Layer 2 loops.

Current EVPN specifications [RFC7432] and [RFC8584] rely on a timer-based approach for transferring the DF role to the newly inserted device. This can cause the following issues:

- \* Loops and duplicates, if the timer value is too short
- \* Prolonged traffic loss, if the timer value is too long

#### 1.4. Design Principles for a Solution

The clock-synchronization solution for fast DF recovery presented in this document follows several design principles and offers multiple advantages, namely:

- \* Complex handshake signaling mechanisms and state machines are avoided in favor of a simple unidirectional signaling approach.
- \* The fast DF recovery solution maintains backwards compatibility (see Section 4) by ensuring that PEs reject any unrecognized new BGP EVPN Extended Community.
- \* Existing DF Election algorithms remain supported.
- \* The fast DF recovery solution is independent of any BGP delays in propagation of Ethernet Segment routes (Route Type 4)
- \* The fast DF recovery solution is agnostic of the actual time synchronization mechanism used; however, an NTP-based representation of time is used for EVPN signaling.

The solution in this document relies on nodes in the topology, more specifically the peering nodes of each Ethernet-Segment, to be clock-synchronized and to advertise the Time Synchronization capability. When this is not the case, or when clocks are badly desynchronized, network convergence and DF Election is no worse than that described in [RFC7432] due to the timestamp range checking (Section 2.2).

## 2. DF Election Synchronization Solution

The fast DF recovery solution relies on the concept of common clock alignment between partner PEs participating in a common Ethernet Segment, i.e., PE1 and PE2 in Figure 1. The main idea is to have all peering PEs of that Ethernet Segment perform DF election and apply the result at the same previously announced time.

The DF Election procedure, as described in [RFC7432] and as optionally signaled in [RFC8584], is applied. All PEs attached to a given Ethernet Segment are clock-synchronized using a networking protocol for clock synchronization (e.g., NTP, Precision Time Protocol (PTP)). Whenever possible, recovery activities for failed PEs SHOULD NOT be initiated until after the underlying clock synchronization protocol has converged to benefit from this document's fast DF recovery procedures. When a new PE is inserted in an Ethernet Segment or when a failed PE of the Ethernet Segment recovers, that PE communicates to peering partners the current time plus the value of the timer for partner discovery from step 2 in Section 8.5 of [RFC7432]. This constitutes an "end time" or "absolute time" as seen from the local PE. That absolute time is called the Service Carving Time (SCT).

A new BGP EVPN Extended Community, the Service Carving Time, is

advertised along with the Ethernet Segment Route Type 4 (RT-4) and communicates the SCT to other partners to ensure an orderly transfer of forwarding duties.

Upon receipt of the new BGP EVPN Extended Community, partner PEs can determine the SCT of the newly inserted PE. To eliminate any potential for duplicate traffic or loops, the concept of "skew" is introduced: a small time offset to ensure a controlled and orderly transition when multiple PE devices are involved. The previously inserted PE(s) must perform service carving first for NDF to DF transitions. The receiving PEs subtract this skew (default = 10 ms) to the Service Carving Time and apply NDF to DF transitions first. This is followed shortly by the NDF to DF transitions on both PEs, after the skew delay. On the recovering PE, all services are already in NDF state, and no skew for DF to NDF transitions is required.

This document proposes a default skew value of 10 ms to allow completion of programming the DF to NDF transitions, but implementations may make the skew larger (or configurable) taking into consideration scale, hardware capabilities, and clock accuracy.

To summarize, all peering PEs perform service carving almost simultaneously at the time announced by the newly added/recovered PE. The newly inserted PE initiates the SCT and triggers service carving immediately on its local timer expiry. The previously inserted PE(s) receiving Ethernet Segment route (RT-4) with an SCT BGP extended community perform service carving shortly before the SCT for DF to NDF transitions and at the SCT for NDF to DF transitions.

## 2.1. BGP Encoding

A BGP extended community, with Type 0x06 and Sub-Type 0x0F, is defined to communicate the SCT for each Ethernet Segment:

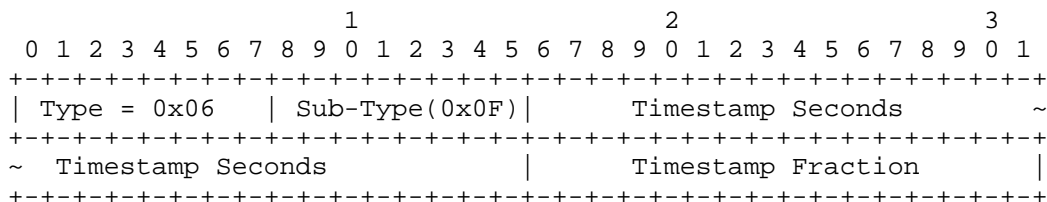


Figure 2: Service Carving Time

The timestamp exchanged uses the NTP prime epoch of 0 h 1 January 1900 UTC [RFC5905] and an adapted form of the 64-bit NTP timestamp format.

The 64-bit NTP timestamp format consists of a 32-bit unsigned seconds field and a 32-bit fraction field, which are encoded in the Service Carving Time as follows:

Timestamp Seconds: 32-bit NTP seconds are encoded in this field.

Timestamp Fraction: The high-order 16 bits of the NTP "Fraction" field are encoded in this field.

When rebuilding a 64-bit NTP timestamp format using the values from a received SCT BGP extended community, the lower-order 16 bits of the NTP "Fraction" field are set to 0. The use of a 16-bit fractional seconds value yields adequate precision of 15 microseconds ( $2^{-16}$  s).

The format of the DF Election Extended Community that is used in this document is:

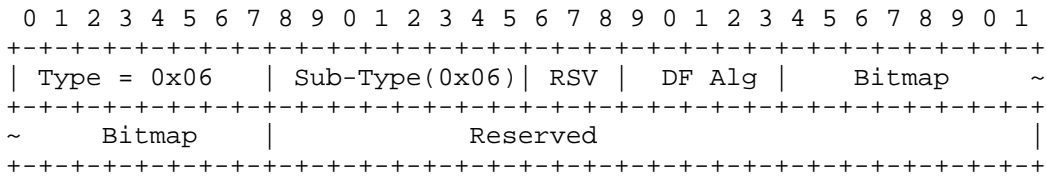


Figure 3: DF Election Extended Community (RFC 8584)

The Bitmap field (2 octets) encodes "capabilities" [RFC8584], where this document introduces a new Time Synchronization capability indicated by "T".

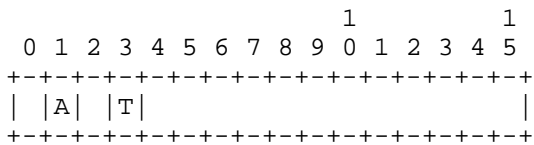


Figure 4: Bitmap Field in the DF Election Extended Community

Bit 3: Time Synchronization (corresponds to Bit 27 of the DF Election Extended Community). When set to 1, it indicates the desire to use the Time Synchronization capability with the rest of the PEs in the Ethernet Segment.

This capability is utilized in conjunction with the agreed-upon DF Election Type. For instance, if all the PE devices in the Ethernet Segment indicate the desire to use the Time Synchronization capability and request the DF Election Type to be the HRW, then the HRW algorithm is used in conjunction with this capability. A PE that does not support the procedures set out in this document or that receives a route from another PE in which the capability is not set MUST NOT delay DF election as this could lead to duplicate traffic in some instances (overlapping DFs).

### 2.2. Timestamp Verification

The NTP Era value is not exchanged, and participating PEs may consider the timestamps to be in the same Era as their local value. A DF Election operation occurring exactly at the next Era transition will be some time on February 7, 2036. Implementors and operators may address credible cases of rollover ambiguity (adjacent Eras n and n+1) as well as the security issue of unreasonably large or unreasonably small NTP timestamps in the following manner.

The procedures in this document address implicitly what occurs with receiving an SCT value in the past. This would be a naturally occurring event with a large BGP propagation delay: the receiving PE treats the DF Election at the peer as having already occurred and proceeds without starting any timer to further delay service carving, effectively falling back on behavior as specified in [RFC7432]. A PE that receives an SCT value smaller than its current time MUST discard the Service Carving Time and SHALL treat the DF Election at the peer as having occurred already.

The more problematic scenario is the PE in Era n+1 that receives an SCT advertised by the PE still in Era n, with a very large SCT value. To address this Era rollover as well as the large values attack vector, implementations MUST validate the received SCT against an upper bound.

It is left to implementations to decide what constitutes an "unreasonably large" SCT value. A recommended approach, however, is to compare the received offset to the local peering timer value. In practice, peering timer values are configured uniformly across

Ethernet Segment peers and may be treated as an upper bound on the offset of received SCT values. A PE that receives an SCT representing an offset larger than the local peering timer MUST discard the SCT and SHALL treat the DF Election at the peer as having already occurred, as above.

### 2.3. Updates to RFC 8584

This document introduces an additional delay to the events and transitions defined for the default DF election algorithm FSM in Section 2.1 of [RFC8584] without changing the FSM state or event definitions themselves.

Upon receiving an RCVD\_ES message, the peering PE's FSM transitions from the DF\_DONE state (indicating the DF election process was complete) to the DF\_CALC state (indicating that a new DF calculation is needed). Due to the SCT included in the Ethernet Segment update, the completion of the DF\_CALC state and the subsequent transition back to the DF\_DONE state are delayed. This delay ensures proper synchronization and prevents conflicts. Consequently, the accompanying forwarding updates to the DF and NDF states are also deferred.

Item 9 in Section 2.1 of [RFC8584], in the list "Corresponding actions when transitions are performed or states are entered/exited", is changed as follows:

- 9. DF\_CALC on CALCULATED: Mark the election result for the VLAN or VLAN bundle.
  - 9.1 If no Service Carving Time is present during the RCVD\_ES event of Action 11, proceed to step 9.4
  - 9.2 If a Service Carving Time is present during the RCVD\_ES event of Action 11, wait until the time indicated by the SCT minus skew before proceeding to step 9.3.
  - 9.3 Assume the role of NDF for the local PE concerning the VLAN or VLAN bundle. Wait the remaining skew time before proceeding to step 9.4.
  - 9.4 Assume the election result's role (DF or NDF) for the local PE concerning the VLAN or VLAN bundle and transition to the DF\_DONE state.

This revised approach ensures proper timing and synchronization in the DF election process, avoiding conflicts and ensuring accurate forwarding updates.

### 3. Synchronization Scenarios

Consider Figure 1 as an example, where initially PE2 has failed and PE1 has taken over. This scenario illustrates the problem with the DF Election mechanism described in Section 8.5 of [RFC7432], specifically in the context of the timer value configured for all PEs on the Ethernet Segment.

The following procedure is based on Section 8.5 of [RFC7432] with the default 3-second timer in step 2.

1. Initial state: PE1 is in a steady-state and PE2 is recovering.
2. Recovery: PE2 recovers at an absolute time of  $t=99$ .
3. Advertisement: PE2 advertises RT-4, sent at  $t=100$ , to its partner (PE1).

4. Timer Start: PE2 starts a 3-second timer to allow the reception of RT-4 from other PE nodes.
5. Immediate carving: PE1 performs service carving immediately upon RT-4 reception, i.e.,  $t=100$  plus some BGP propagation delay.
6. Delayed Carving: PE2 performs service carving at time  $t=103$ .

[RFC7432] favors traffic drops over duplicate traffic. With the above procedure, traffic drops will occur as part of each PE recovery sequence since PE1 transitions some VLANs to an NDF immediately upon RT-4 reception. The timer value (default = 3 seconds) directly affects the duration of the packet drops. A shorter (or zero) timer may result in duplicate traffic or traffic loops.

The following procedure is based on the SCT approach:

1. Initial state: PE1 is in a steady state, and PE2 is recovering.
2. Recovery: PE2 recovers at an absolute time of  $t=99$ .
3. Timer Start: PE2 starts at  $t=100$  a 3-second timer to allow the reception of RT-4 from other PE nodes.
4. Advertisement: PE2 advertises RT-4, sent at  $t=100$ , with a target SCT value of  $t=103$  to its partner (PE1).
5. Service Carving Timer: PE1 starts the service carving timer, with the remaining time until  $t=103$ .
6. Simultaneous Carving: Both PE1 and PE2 carve at an absolute time of  $t=103$ .

To maintain the preference for minimal loss over duplicate traffic, PE1 SHOULD carve slightly before PE2 (with skew). The recovering PE2 performs both DF-to-NDF and NDF-to-DF transitions per VLAN at the timer's expiry. The original PE1, which received the SCT, applies the following:

- \* DF-to-NDF Transition(s): at  $t=\text{SCT} - \text{skew}$ , where both PEs are NDF for the skew duration.
- \* NDF-to-DF Transition(s): at  $t=\text{SCT}$ .

This split behavior ensures a smooth DF role transition with minimal loss.

The SCT approach mitigates the negative effect of requiring a timer for discovery of Ethernet Segment (ES) RT-4 from other PE nodes. Furthermore, the BGP transmission delay (from PE2 to PE1) of the ES RT-4 becomes a non-issue. The SCT approach shortens the 3-second timer window to the order of milliseconds.

The peering timer is a configurable value where 3 seconds represents the default. Configuring a timer value of 0, or so small as to expire during propagation of the BGP routes, is outside the scope of this document. In reality, the use of the SCT approach presented in this document encourages the use of larger peering timer values to overcome any sort of BGP route propagation delays.

### 3.1. Concurrent Recoveries

In the eventuality that two or more PEs in a peering Ethernet Segment group are recovering concurrently or roughly at the same time, each will advertise a SCT. This SCT value would correspond to what each

recovering PE considers the "end time" for DF Election. A similar situation arises in sequentially recovering PEs, when a second PE recovers approximately at the time of the first PE's advertised SCT expiry and with its own new SCT-2 outside of the initial SCT window.

In the case of multiple concurrent DF elections, each initiated by one of the recovering PEs, the SCTs must be ordered chronologically. All PEs SHALL execute only a single DF Election at the service carving time corresponding to the largest (latest) received timestamp value. This DF Election will lead peering PEs into a single coordinated DF Election update.

Example:

1. Initial State: PE1 is in a steady state, with services elected at PE1.
2. Recovery of PE2: PE2 recovers at time  $t=100$  and advertises RT-4 with a target SCT value of  $t=103$  to its partner (PE1).
3. Timer Initiation by PE2: PE2 starts a 3-second timer to allow the reception of RT-4 from other PE nodes.
4. Timer Initiation by PE1: PE1 starts the service carving timer, with the remaining time until  $t=103$ .
5. Recovery of PE3: PE3 recovers at time  $t=102$  and advertises RT-4 with a target SCT value of  $t=105$  to its partners (PE1, PE2).
6. Timer Initiation by PE3: PE3 starts a 3-second timer to allow the reception of RT-4 from other PE nodes.
7. Timer Update by PE2: PE2 cancels the running timer and starts the service carving timer with the remaining time until  $t=105$ .
8. Timer Update by PE1: PE1 updates its service carving timer, with the remaining time until  $t=105$ .
9. Service Carving: PE1, PE2, and PE3 perform service carving at the absolute time of  $t=105$ .

In the eventuality that a PE in an Ethernet Segment group recovers during the discovery window specified in Section 8.5 of [RFC7432] and does not support or advertise the T-bit, all PEs in the current peering sequence SHALL immediately revert to the default behavior described in [RFC7432].

#### 4. Backwards Compatibility

For the DF election procedures to achieve global convergence and unanimity within a redundancy group, it is essential that all participating PEs agree on the DF election algorithm to be employed. However, it is possible that some PEs may continue to use the existing modulo-based DF election algorithm from [RFC7432] and not utilize the new SCT BGP extended community. PEs that operate using the baseline DF election mechanism will simply discard the new SCT BGP extended community as unrecognized.

A PE can indicate its willingness to support clock-synchronized carving by signaling the new "T" DF Election Capability and including the new SCT BGP extended community along with the Ethernet Segment Route Type 4. If one or more PEs attached to the Ethernet Segment do not signal  $T=1$ , then all PEs in the Ethernet Segment SHALL revert to the timer-based approach as specified in [RFC7432]. This reversion is particularly crucial in preventing VLAN shuffling when more than two PEs are involved.

In the event a new or extra RT-4 is received without the new "T" DF Election Capability in the midst of an ongoing DF Election sequence, all SCT-based delays are canceled, and the DF Election is immediately applied as specified in [RFC7432], as if no SCT had been previously exchanged.

## 5. Security Considerations

The mechanisms in this document use the EVPN control plane as defined in [RFC7432]. Security considerations described in [RFC7432] are equally applicable.

For the new SCT Extended Community, attack vectors may be setting the value to zero, to a value in the past, or to large times in the future. Handling of this attack vector is addressed in Section 2.2 alongside NTP Era rollover ambiguity.

This document uses MPLS- and IP-based tunnel technologies to support data plane transport. Security considerations described in [RFC7432] and [RFC8365] are equally applicable.

## 6. IANA Considerations

IANA has made the following assignment in the "EVPN Extended Community Sub-Types" registry set up by [RFC7153].

Sub-Type Value	Name	Reference
0x0F	Service Carving Time	RFC 9722

Table 1

IANA has made the following assignment in the "DF Election Capabilities" registry set up by [RFC8584].

Bit	Name	Reference
3	Time Synchronization	RFC 9722

Table 2

## 7. References

### 7.1. 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, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC5905] Mills, D., Martin, J., Ed., Burbank, J., and W. Kasch, "Network Time Protocol Version 4: Protocol and Algorithms Specification", RFC 5905, DOI 10.17487/RFC5905, June 2010, <<https://www.rfc-editor.org/info/rfc5905>>.
- [RFC7153] Rosen, E. and Y. Rekhter, "IANA Registries for BGP Extended Communities", RFC 7153, DOI 10.17487/RFC7153, March 2014, <<https://www.rfc-editor.org/info/rfc7153>>.
- [RFC7432] Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based

Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <<https://www.rfc-editor.org/info/rfc7432>>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

[RFC8365] Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <<https://www.rfc-editor.org/info/rfc8365>>.

[RFC8584] Rabadan, J., Ed., Mohanty, S., Ed., Sajassi, A., Drake, J., Nagaraj, K., and S. Sathappan, "Framework for Ethernet VPN Designated Forwarder Election Extensibility", RFC 8584, DOI 10.17487/RFC8584, April 2019, <<https://www.rfc-editor.org/info/rfc8584>>.

## 7.2. Informative References

[HRW98] Thaler, D. and C. Ravishankar, "Using Name-Based Mappings to Increase Hit Rates", IEEE/ACM Transactions on Networking, vol. 6, no. 1, February 1998, <<https://www.microsoft.com/en-us/research/wp-content/uploads/2017/02/HRW98.pdf>>.

## Acknowledgements

Authors would like to acknowledge helpful comments and contributions of Satya Mohanty and Bharath Vasudevan. Also thank you to Anoop Ghanwani and Gunter van de Velde for their thorough review with valuable comments and corrections.

## Contributors

In addition to the authors listed on the front page, the following coauthors have also contributed substantially to this document:

Gaurav Badoni  
Cisco  
Email: [gbadoni@cisco.com](mailto:gbadoni@cisco.com)

Dhananjaya Rao  
Cisco  
Email: [dhrao@cisco.com](mailto:dhrao@cisco.com)

## Authors' Addresses

Patrice Brissette  
Cisco  
Email: [pbrisset@cisco.com](mailto:pbrisset@cisco.com)

Ali Sajassi  
Cisco  
Email: [sajassi@cisco.com](mailto:sajassi@cisco.com)

Luc Andr Burdet (editor)  
Cisco  
Email: [lburdet@cisco.com](mailto:lburdet@cisco.com)

John Drake  
Independent  
Email: je\_drake@yahoo.com

Jorge Rabadan  
Nokia  
Email: jorge.rabadan@nokia.com