

Detnet Working Group  
Internet-Draft  
Intended status: Standards Track  
Expires: September 6, 2018

S. Bryant  
M. Chen  
Huawei Technologies  
March 05, 2018

Operation of Deterministic Networks over MPLS  
draft-bryant-detnet-mpls-dp-00

## Abstract

This document specifies Deterministic Networking data plane operation over an MPLS Packet Switched Networks.

This document is a derivative work from [draft-ietf-detnet-dp-sol-01](#).

Whether this is published as a stand-alone text, or serves as a focal point to refine the MPLS design and the key point are subsequently re-merged with [draft-ietf-detnet-dp-sol-01](#) is a matter for the DETNET WG, as is the editorship of any WG text.

## 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 6, 2018.

## Copyright Notice

Copyright (c) 2018 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

Internet-Draft

Deterministic MPLS

March 2018

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

<a href="#">1.</a>	<a href="#">Introduction . . . . .</a>	<a href="#">3</a>
<a href="#">2.</a>	<a href="#">Terminology . . . . .</a>	<a href="#">3</a>
<a href="#">2.1.</a>	<a href="#">Terms used in this document . . . . .</a>	<a href="#">3</a>
<a href="#">2.2.</a>	<a href="#">Abbreviations . . . . .</a>	<a href="#">4</a>
<a href="#">3.</a>	<a href="#">Requirements language . . . . .</a>	<a href="#">5</a>
<a href="#">4.</a>	<a href="#">DetNet Over an MPLS Underlay . . . . .</a>	<a href="#">5</a>
<a href="#">5.</a>	<a href="#">DetNet over MPLS Encapsulation Components . . . . .</a>	<a href="#">8</a>
<a href="#">5.1.</a>	<a href="#">Basic Data Plane Encapsulation . . . . .</a>	<a href="#">8</a>
<a href="#">5.2.</a>	<a href="#">DetNet Control Word . . . . .</a>	<a href="#">9</a>
<a href="#">5.3.</a>	<a href="#">Flow identification . . . . .</a>	<a href="#">10</a>
<a href="#">5.4.</a>	<a href="#">OAM Indication . . . . .</a>	<a href="#">11</a>
<a href="#">5.5.</a>	<a href="#">Flow Aggregation . . . . .</a>	<a href="#">11</a>
<a href="#">5.6.</a>	<a href="#">Aggregation at the LSP . . . . .</a>	<a href="#">12</a>
<a href="#">5.7.</a>	<a href="#">Aggregating DetNet flows as a new DetNet flow . . . . .</a>	<a href="#">12</a>
<a href="#">6.</a>	<a href="#">Simple Aggregation at the DetNet layer . . . . .</a>	<a href="#">13</a>
<a href="#">7.</a>	<a href="#">Indication of the DetNet Payload Type. . . . .</a>	<a href="#">13</a>
<a href="#">8.</a>	<a href="#">Operation of the PREF Functions . . . . .</a>	<a href="#">14</a>
<a href="#">8.1.</a>	<a href="#">Operation of PR . . . . .</a>	<a href="#">14</a>
<a href="#">8.2.</a>	<a href="#">Operation of EF . . . . .</a>	<a href="#">15</a>
<a href="#">8.3.</a>	<a href="#">Packet reordering considerations . . . . .</a>	<a href="#">15</a>
<a href="#">8.4.</a>	<a href="#">Indication of PREF processing . . . . .</a>	<a href="#">16</a>
<a href="#">8.5.</a>	<a href="#">Placement of PR and EF in a node . . . . .</a>	<a href="#">16</a>
<a href="#">9.</a>	<a href="#">Operation at DetNet Node types . . . . .</a>	<a href="#">17</a>
<a href="#">9.1.</a>	<a href="#">Operation at an Ingress Edge (PE) Node . . . . .</a>	<a href="#">17</a>
<a href="#">9.2.</a>	<a href="#">Operation at a Relay node (S-PE) . . . . .</a>	<a href="#">17</a>
<a href="#">9.3.</a>	<a href="#">Operation at an Egress Edge (PE) Node . . . . .</a>	<a href="#">18</a>
<a href="#">9.4.</a>	<a href="#">Operation at a Transit(P) Node . . . . .</a>	<a href="#">18</a>
<a href="#">10.</a>	<a href="#">Other DetNet data plane considerations . . . . .</a>	<a href="#">19</a>
<a href="#">10.1.</a>	<a href="#">Class of Service . . . . .</a>	<a href="#">19</a>
<a href="#">10.2.</a>	<a href="#">Quality of Service . . . . .</a>	<a href="#">19</a>
<a href="#">10.3.</a>	<a href="#">Bidirectional traffic . . . . .</a>	<a href="#">20</a>
<a href="#">10.4.</a>	<a href="#">Layer 2 addressing and QoS Considerations . . . . .</a>	<a href="#">21</a>
<a href="#">11.</a>	<a href="#">Management and control considerations . . . . .</a>	<a href="#">22</a>
<a href="#">11.1.</a>	<a href="#">S-Label assignment and distribution . . . . .</a>	<a href="#">22</a>
<a href="#">11.2.</a>	<a href="#">Explicit routes . . . . .</a>	<a href="#">22</a>
<a href="#">12.</a>	<a href="#">Security considerations . . . . .</a>	<a href="#">22</a>

<a href="#">13.</a>	IANA considerations . . . . .	<a href="#">23</a>
<a href="#">13.1.</a>	Acknowledgments . . . . .	<a href="#">23</a>
<a href="#">14.</a>	References . . . . .	<a href="#">23</a>
<a href="#">14.1.</a>	Normative References . . . . .	<a href="#">23</a>
<a href="#">14.2.</a>	Informative References . . . . .	<a href="#">23</a>

Authors' Addresses . . . . .	<a href="#">26</a>
------------------------------	--------------------

## [1.](#) Introduction

This document is a derivative work from [[draft-ietf-detnet-dp-sol-01](#)].

Editor's Note: We need to point to the exact version that this was derived from, not a generic version of [[I-D.ietf-detnet-dp-sol](#)].

Whether this is published as a stand-alone text, or serves as a focal point to refine the MPLS design and the key point are subsequently re-merged with [draft-ietf-detnet-dp-sol-01](#) is a matter for the DETNET WG.

Deterministic Networking (DetNet) is a service that can be offered by a network to DetNet flows. DetNet provides these flows extremely low packet loss rates and assured maximum end-to-end delivery latency. General background and concepts of DetNet can be found in [[I-D.ietf-detnet-architecture](#)].

This document specifies the encapsulation and operation of deterministic networking over an MPLS data-plane. The approach is modeled on the operation of Pseudowires (PW) over an MPLS Packet Switched Network (PSN) [[RFC3985](#)] [RFC4385].

The DetNet transport layer functionality that provides congestion protection for DetNet flows is assumed to be in place in a DetNet node.

This document does not define the associated control plane functions, or Operations, Administration, and Maintenance (OAM). It also does not specify traffic handling capabilities required to deliver congestion protection and latency control for DetNet flows at the DetNet transport layer.

## [2.](#) Terminology

### [2.1.](#) Terms used in this document

Editor's note: This section needs to be reviewed when the body of the text is closer to completion.

This document uses the terminology established in the DetNet architecture [[I-D.ietf-detnet-architecture](#)] and the DetNet Data Plane Solution Alternatives [[I-D.ietf-detnet-dp-alt](#)].

**T-Label** A label used to identify the LSP used to transport a DetNet flow across an MPLS PSN, e.g., a hop-by-hop label used between label switching routers (LSR).

**S-Label** A DetNet "service" label that is used between DetNet nodes that implement also the DetNet service layer functions. An S-Label is also used to identify a DetNet flow at DetNet service layer.

**Local-ID** A DetNet Edge and Relay node internal construct that uniquely identifies a DetNet flow within a node and never appear on-wire. It may be used to select proper forwarding and/or DetNet specific service function.

**PREF** A Packet Replication and Elimination Function (PREF) does the replication and elimination processing of DetNet flow packets in edge or relay nodes. The replication function is essentially the existing 1+1 protection mechanism. The elimination function reuses and extends the existing duplicate detection mechanism to operate over multiple (separate) DetNet member flows of a DetNet compound flow.

**DetNet Control Word** A control word used for sequencing and identifying duplicate packets at the DetNet service layer.

### [2.2.](#) Abbreviations

Editor's note: This section needs to be reviewed when the body of the text is closer to completion.

The following abbreviations used in this document:

AC Attachment Circuit.

CE Customer Edge equipment.

CoS Class of Service.

CW Control Word.

d-CW DetNet Control Word.

DetNet Deterministic Networking.

DF DetNet Flow.

L2VPN Layer 2 Virtual Private Network.

LSR Label Switching Router.

MPLS Multiprotocol Label Switching.

MPLS-TP Multiprotocol Label Switching - Transport Profile.

MS-PW Multi-Segment Pseudowire (MS-PW).

NSP Native Service Processing.

OAM Operations, Administration, and Maintenance.

PE Provider Edge.

PREF Packet Replication and Elimination Function.

PSN Packet Switched Network.

PW Pseudowire.

QoS Quality of Service.

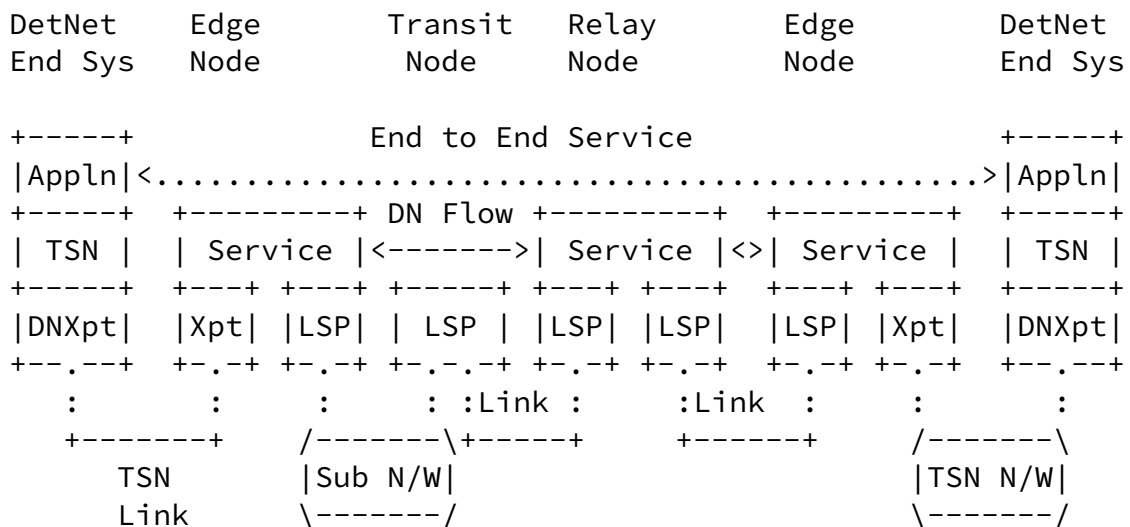
TSN Time-Sensitive Network.

### 3. Requirements language

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 [[RFC2119](#)].

### 4. DetNet Over an MPLS Underlay

Figure 1 Shows the basic components of a DetNet enabled MPLS network used to transport TSN traffic using an MPLS transport.



LSP = MPLS Transport  
DNXpt & Xpt = DetNet Transport

Figure 1: A Simple DetNet Enabled MPLS Network

TSN End Systems send and receive packets over the DetNet. The supported packet types are IP and Ethernet.

Edge Nodes are responsible for the imposition and disposition of the required DetNet encapsulation. These are functionally similar to pseudowire (PW) Terminating Provide Edge (T-PE) equipment.

Relay nodes are strategically placed and used enhance the reliability of delivery by enabling the replication of packets so that multiple copies, possibly over multiple paths. They also reduce the impact of replication by the eliminating surplus copies of DetNet packets. Replication and elimination may also be performed at ingress and egress edge nodes respectively.

Edge nodes, and relay nodes are aware of the needs of particular DetNet flows and take care to process them in accordance with the required performance needs.

Transit nodes are normal MPLS Label Switching Routers (LSRs). They are unaware of the special requirements of DetNet flows, although they may be configured to provide traffic engineering services to them to enhance the prospect of them meeting the required service level agreement (SLA).

The MPLS LSP may be provided by any MPLS method (LSP, RSVP-TE, MPLS-TP, or MPLS Segment Routing (SR)).

Figure 2 illustrates how end to end MPLS-based DetNet service is provided in a more detail.

In this case, the end systems are able to send and receive DetNet flows. For example, an end system sends data encapsulated in MPLS. The 'X' in the end systems, edge and relay nodes represents potential DetNet flow packet replication and elimination points. Here the relay nodes may change the underlying transport, for example tunneling MPLS over IP [[draft-xu-mpls-sr-over-ip](#)], or simply

interconnect network segments.

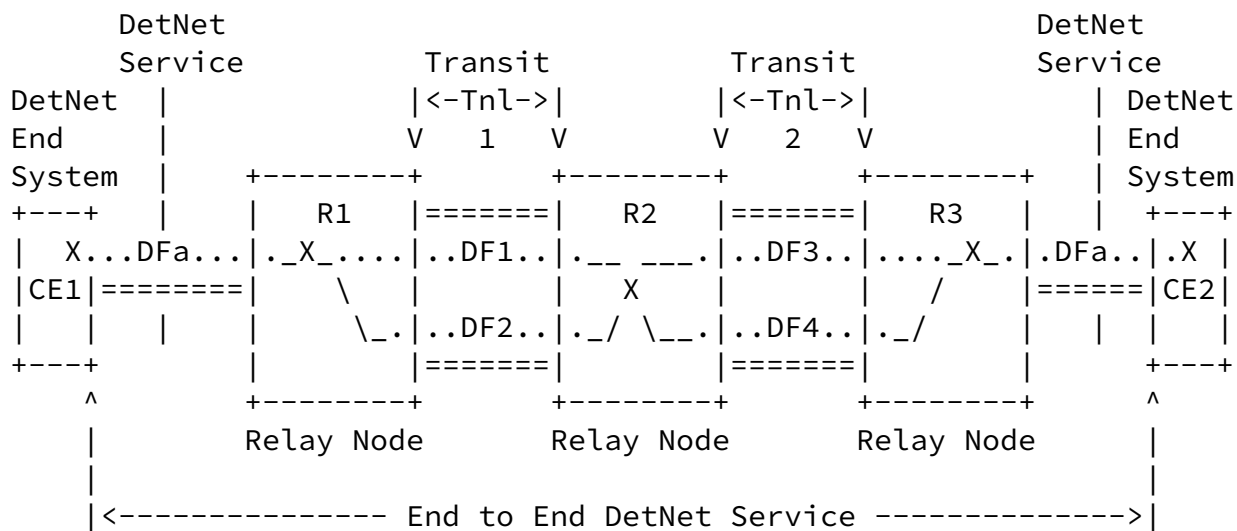


Figure 2: Flows in a DetNet Enabled MPLS Network

An example MPLS DetNet network fragment and packet flow is illustrated in Figure 3.

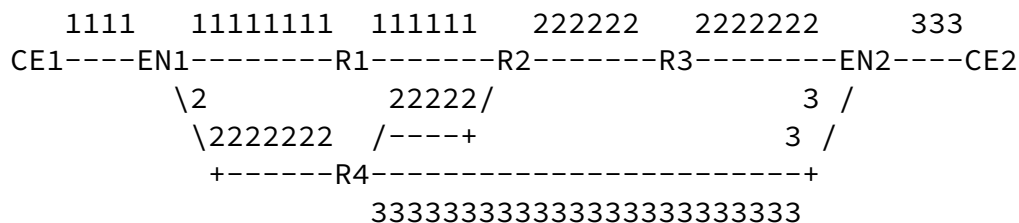


Figure 3: Example Packet flow in DetNet Enabled MPLS Network

Customer Equipment CE1 send a packet into the DetNet enabled MPLS network. Edge Node EN1 makes a copy of the packet and encapsulates each copy as a DetNet packet (packet 1 and packet 2). It sends one copy (1) to Relay Node R1 and the other copy (2) to Relay Node R4. R1 sends packet copy 1 to R2. R4 sends one copy to R2, and a further copy (3) to EN2. R2 receives copy (2) before copy 1, and so eliminated copy (1) sending only (2) to EN2. EN2 receives copy (3) first sending it to CE2 and eliminating copy (2). Note the number

illustrates the replication instance and should not be confused with



the sequence number which remains unchanged in all packet copies.

The above is of course illustrative of many network scenarios that can be configured. Between a pair of relay nodes there may be one or more transport nodes that simply forward the DetNet traffic, but these are omitted for clarity.

## 5. DetNet over MPLS Encapsulation Components

To carry DetNet over MPLS the following is required:

1. A method of identifying the MPLS payload type.
2. A method of identifying the DetNet flow group to the processing element.
3. A method of distinguishing DetNet OAM packets from DetNet data packets.
4. A method of carrying the DetNet sequence number.
5. A suitable LSP to deliver the packet to the egress PE.
6. A method of carrying queuing and forwarding indication.

In this design an MPLS service label (the S-Label), similar to a pseudowire (PW) label [[RFC3985](#)], is used to identify both the DetNet flow identity and the payload MPLS payload type satisfying (1) and (2) in the list above. OAM traffic discrimination happens through the use of the Associated Channel method described in [[RFC4385](#)]. The sequence number is carried in the DetNet Control word which carries the Data/OAM discriminator. The LSP used to transport the DetNet packet may be of any type (MPLS-LDP, MPLS-TE, MPLS-TP[[RFC5921](#)], or MPLS-SR[I-D.ietf-spring-segment-routing-mpls]). The LSP (T-Label) label and/or the S-Label may be used to indicate the queue processing as well as the forwarding parameters.

Note that when the network consists only of DetNet enabled nodes with no aggregation, Penultimate Hop Popping (PHP) means that the only label in the label stack is the S-label.

### 5.1. Basic Data Plane Encapsulation

Figure 4 shows a DetNet data plane MPLS encapsulation. This is modeled on the encapsulation of pseudowires over MPLS [[RFC3985](#)].

The encapsulation consists of:

- o DetNet control word (d-CW) containing sequencing information for packet replication and duplicate elimination purposes, and the OAM indicator. There MUST a separate sequence number space for each DetNet flow.
- o DetNet Label (S-label) that identifies a DetNet flow to the peer node that is to process it. The S-Label is allocated from the platform label space.
- o Zero or more MPLS transport LSP label(s) (T-label) used to direct the packet along the label switched path (LSP) to the next peer node along the path. When Penultimate Hop Popping is in use there will be no label T-label in the protocol stack on the final hop.
- o The necessary data-link encapsulation is then applied prior to transmission over the physical media.

RFC Editor - if you ever get this text please remove this para (text to make compiler work).

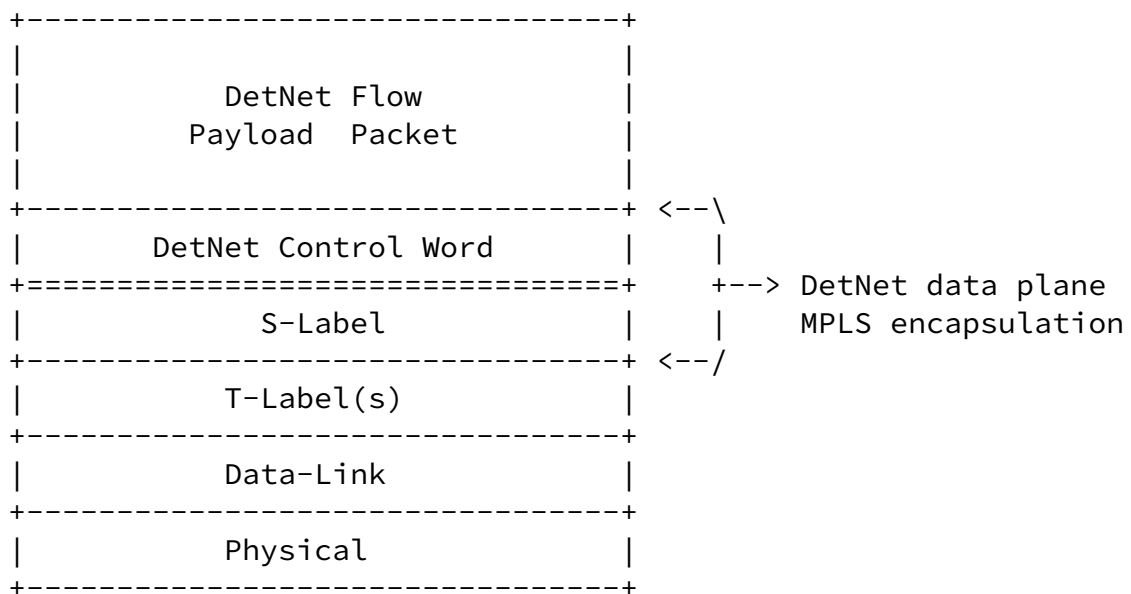


Figure 4: MPLS Encapsulation of DetNet

Flow aggregation may be necessary to achieve the required scaling. The extensions to basic encapsulation needed to support flow aggregation are described in [Section 5.5](#).

## 5.2. DetNet Control Word

A DetNet Control Word (d-CW) conforms to the Generic PW MPLS Control

Word (PWWCW) defined in [[RFC4385](#)] and is shown in Figure 5. In a DetNet data packet the upper nibble of the d-CW MUST be set to zero

(0). The effective sequence number bit length is between 0 and 28 bits, and configured either by a control plane or manually for each DetNet flow. The sequence number is aligned to the right (least significant bits) and unused bits MUST be set to zero (0). Each DetNet flow MUST have its own sequence number counter. The sequence number is incremented by one for each new packet.

Editor's note: We need to consider whether it is better to allow a multiplicity of sequence number lengths with a length configured for each flow, or a uniform sequence number of 28. On reflection it seems better to go for the simplicity of a standard length. If for any reason a different length becomes desirable, then it is relatively simple to define another type of DetNet d-CW with a different standard sequence number length and there is no ambiguity of operation since the sequence number length is a parameter of the S-Label.

The d-CW MUST always be present in a packet. In cases where the sequence number is not used (e.g., for DetNet-t-flows) the control plane or the manual configuration has to define zero (0) bit length sequence number and the value of the sequence number MUST be set to zero (0).

Editors Note: Do we set length zero or say it is not used? Also need to add text to stop relay nodes and egress edge nodes from processing the s/n otherwise only one packet ever gets through!

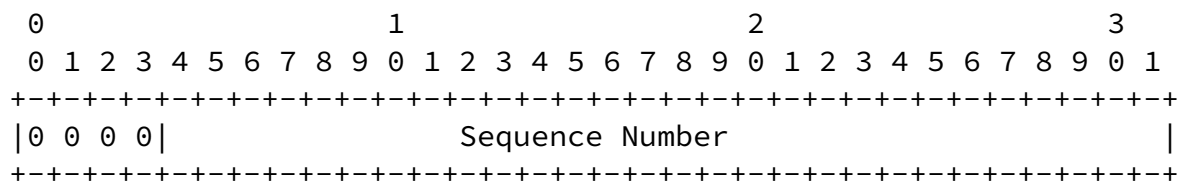


Figure 5: DetNet Control Word

### 5.3. Flow identification

DetNet flow identification at a DetNet service layer is realized by an S-label. It maps a DetNet flow to a specific d-CW in a DetNet

node.

For a data flow the S-label used for flow identification MUST be bottom label of the label stack for a DetNet-s- or DetNet-st-flow and MUST precede the d-CW.

Editor's note: We have specified the above for \_data\_ to leave the door open for the GAL based OAM method as an alternative to the ACH mechanism currently specified. When using GAL the GAL label would be

after the S-Label. Do we leave this option in or shut the door on it?

An S-label for a single DetNet flow MUST be unique at the peer at the node that is to process it. The S-label is stored in the platform label table to allow for DetNet packet processing independent of the interface on which the packet is received on.

#### [5.4.](#) OAM Indication

OAM follows the procedures set out in [[RFC5085](#)] with the restriction that only Virtual Circuit Connectivity Verification (VCCV) type 1 is supported.

Editor's Note - is this restriction acceptable? Do we need to support GAL based OAM indication?

As shown in Figure 3 of [[RFC5085](#)] when the first nibble of the d-CW is 0x0 the payload following the d-CW is normal user data. However, when the first nibble of the d-CW is 0x1, the payload that follows the d-DW is an OAM payload with the OAM type indicated by the value in the d-CW Channel Type field.

The reader is referred to [[RFC5085](#)] for a more detailed description of the Associated Channel mechanism, and to the DetNet work on OAM for more information DetNet OAM.

#### [5.5.](#) Flow Aggregation

The ability to aggregate individual flows, and their associated resource control, into a larger aggregate is an important technique for improving scaling of control in the data, management and control

planes. The DetNet data plane allows for the aggregation of DetNet flows, to improved scaling. There are three methods of introducing flow aggregation:

1. Aggregate at the LSP (Transport)
2. Aggregating DetNet flows as a new DetNet flow
3. Simple Aggregation at the DetNet layer

A further method of using SR to perform aggregation is for further study.

The resource control and management aspects of aggregation (including the queuing/shaping/ policing implications) will be covered in other documents.

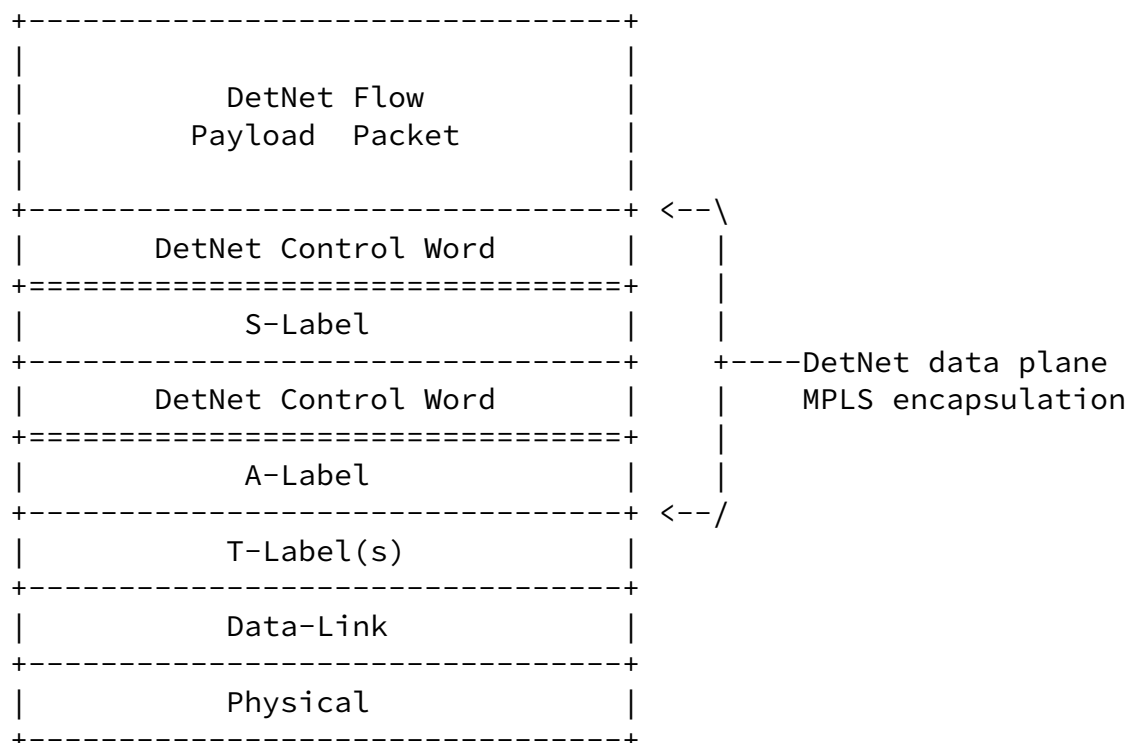
#### [5.6.](#) Aggregation at the LSP

DetNet flows transported via MPLS can leverage MPLS-TE's existing support for hierarchical LSPs (H-LSPs), see [[RFC4206](#)]. H-LSPs are typically used to aggregate control and resources, they may also be used to provide OAM or protection for the aggregated LSPs. Arbitrary levels of aggregation naturally falls out of the definition for hierarchy and the MPLS label stack [[RFC3032](#)]. DetNet nodes which support aggregation (LSP hierarchy) map one or more LSPs (labels) into and from an H-LSP. Both carried LSPs and H-LSPs may or may not use the TC field, i.e., L-LSPs or E-LSPs. Such nodes will need to ensure that traffic from aggregated LSPs are placed (shaped/policed/enqueued) onto the H-LSPs in a fashion that ensures the required DetNet service is preserved.

Additional details of the traffic control capabilities needed at a DetNet-aware node may be covered in the new service descriptions mentioned above or in separate future documents. Management and control plane mechanisms will also need to ensure that the service required on the aggregate flow (H-LSP or DSCP) are provided, which may include the discarding or remarking mentioned in the previous sections.

#### [5.7.](#) Aggregating DetNet flows as a new DetNet flow

An aggregate can be built by layering DetNet flows as shown below:

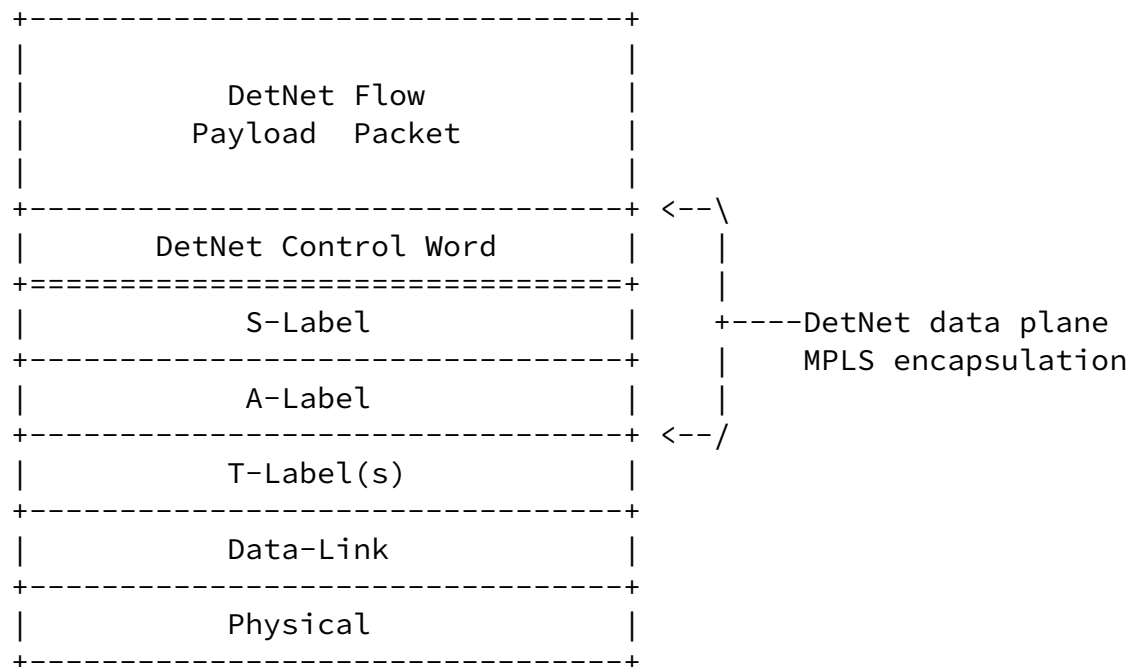


Both the Aggregation (A) label and the S-label have their MPLS S bit set indicating bottom of stack, and the d-CW allows the PREF functions to work.

It is a property of the A-label that what follows is d-CW followed by an S-label. A relay node processing the A-label would not know the underlying payload type. This would only be known to a node that was a peer of the node imposing the S-label. However there is no real need for it to know the payload type during aggregation processing.

## 6. Simple Aggregation at the DetNet layer

Another approach would be not to include a d-CW for the aggregated flow. This would be functionally similar to aggregation at the transport layer using H-LSPs, but would confine knowledge of the aggregation to the DetNet layer. Such an approach shares the disadvantage that PREF operations would not be possible. OAM operation in this mode is for further study.



## 7. Indication of the DetNet Payload Type.

The only node types that needs to know the payload type is the ingress node which has to know how to process the packet it receives on the ingress AC, and egress edge node which has to know how to prepare the packet for transmission on the egress AC.

On ingress a DetNet edge node has to classify the packets into those that are for transmission as Detnet packets and those that are for transmission as "normal" packets at one of more lower priorities.

The packet type is indicated to the egress edge node through the value of the S-label. Thus when the egress edge node looks up the S-label one of the parameters returned is the packet type which in turn tells the egress edge node how to prepare the packet for transmission over the egress AC.

Editor's note: Do we only find one type on the ingress and egress or do we need to run missed mode, i.e. will we have to send some packets across the network as Ethernet packets and some as IP packets?

## 8. Operation of the PREF Functions

The Packet Replication and Elimination Functions (PREF) are designed to enhance the reliability of delivery by network layer packet replication (PR), whilst at the same time minimizing network congestion and the duplicate delivery of packets over an egress AC by eliminating duplicate packets (EF).

PR and EF are independent functions that operate on a DetNet flow at strategic places in the network. The placement of a function is a matter for the network operator.

### [8.1.](#) Operation of PR

A PR function creates two or more copies of a packet, and forwards a copy to each of the designated peers of the replicating node. A PR function may be placed in an ingress edge node, a relay node or an egress edge node.

We consider first a DetNet relay node. The packet is received from the upstream DetNet peer, and if present the T-label is popped. The S-label is looked up in the platform label table and if the forwarding instructions indicate that replication is required, the following happens for each next hop in the DetNet layer:

- o A copy of the payload is made.
- o An identical copy of the d-CW from the original packet is pushed.
- o The S-Label required by the next hop in the DetNet layer is pushed.
- o The DetNet packet copy is forwarded on the LSP to the next hop in the DetNet layer using normal MPLS forwarding.

An ingress node operates in the same way as a relay node, except that it is responsible for initial encapsulation of the packet. The packet is received from the AC, classified and prepared for

forwarding over the DetNet network as described in [Section 9.1](#), except that for each next-hop in the DetNet Layer (i.e. each node that is to receive a copy of the DetNet packet) :



- o A copy of the payload is made.
- o The d-CW is constructed using the next sequence number in the sequence associated with this service.
- o An identical copy of the d-CW is pushed.
- o The S-Label required by the next hop in the DetNet layer to recognize this service is pushed.
- o The DetNet packet copy is forwarded on the LSP to the next hop in the DetNet layer using normal MPLS forwarding.

An egress node operates in the same way as as described in [Section 9.1](#), except that where the S-Label indicates that the the packet is to be forwarded to a multiply attached system in the payload layer a similar copy (modified as needed to conform to any MAC requirements) is forwarded out of each egress AC.

Editor's Note: For normal PW, there will be one to one AC to PW binding relationship, for DetNet, no matter the ingress or egress side, there may be multiple AC corresponding to a single DetNet PW (S-label), should we state explicitly?

## [8.2.](#) Operation of EF

The EF function eliminates duplicate copies of a packet. The node identifies the service from the packet S-Label. If the S-Label indicates that the packet elimination function is required to operate on this service at this node, it uses the sequence number in the d-CW to determine whether or not the packet is a duplicate that must be eliminated, and precedes accordingly.

The EF may be placed in a Relay node, or an Edge Egress node.

## [8.3.](#) Packet reordering considerations

The DetNet service layer introduces packet reordering functionality for use in DetNet edge and relay node and end system packet processing. The reordering functionality MAY be enabled in a DetNet node. The reordering functionality relies on a presence of sequence numbers the d-CW.

#### [8.4.](#) Indication of PREF processing

The indication that PR or EF processing is needed at a DetNet Relay node or a DetNet egress edge node is carried as an implicit characteristic of the S-Label. Thus, when a service is established the ingress edge node is configured to use an active rather than a static zero sequence number, and DetNet relays and egress edge nodes are configured to run PR and/or EF functionality on on services identified by specific S-labels.

#### [8.5.](#) Placement of PR and EF in a node

This section is not normative.

The placement of the PR and EF functions within the node is a matter for the node designers, and this specification makes no determination on this matter. However the placement may well have implications for the management and control of a node, and thus the following is worth noting.

In a bladed system a common processing model is to analyze the packet for forwarding close to the ingress interface and to either to fully prepare it for forwarding as part of ingress processing, or to package it with internal metadata for final preparation close to the egress interface. In such systems the natural place to perform replication is as part of the ingress processing since egress processors often do not have the capability (for example processing power) to further process the packet, often do not normally have the required data paths to facilitate replication for transmission to other line cards.

Furthermore, in bladed systems it is to be expected that a packet from a peer in the MPLS layer may arrive on any of the blades. Whilst in principle some constraints could be applied on which interfaces the packets will arrive on, experience shows that such constraints are operationally impractical. To perform elimination state must be shared amongst the forwarders performing elimination. Sharing the state required for elimination between forwarders on different blades without sacrificing performance is technically difficult. Thus, the natural place for elimination in a distributed design is close to the egress interface.

On mono-processor systems these constraints do not apply in quite the same way, although even in this case it is necessary for the equipment designer to consider the implications of particular forwarder design, for example the allocation of Network Processing Unit (NPU) cores to particular interfaces.

A bladed system may place the PR and EF functions on a processing function other than the ingress or egress forwarding processor, whereupon the mono-processor considerations apply.

As noted above the placement of the PR and EF functions is a matter for the system designer. However it is important that nothing in the control, configuration, or OAM system results in undue difficulties for any of the forwarding models.

## [9.](#) Operation at DetNet Node types

This section considers the operations at the various DetNet node types in more detail.

### [9.1.](#) Operation at an Ingress Edge (PE) Node

An ingress Edge Node first classifies received traffic into DetNet flows and non-Detnet traffic. The DetNet flows are sent over the DetNet network using the procedures defined in the document. The classification process and the handling of non-DetNet traffic is out of scope for this document.

The processing of non-DetNet flows is currently outside the scope of this document

The packet is encapsulated as described in [Section 5.1](#). First, if the flow is to be carried as IP the MAC header and checksum are removed. Otherwise the checksum is removed. The d-CW is constructed using the next sequence number associated with the service, and is pushed. The S-label corresponding to this DetNet flow is then pushed. The packet is then forwarded to the required egress edge-node by pushing the required T-Labels and the required data-link headers.

Editor's Note: we need text on how we indicate IPv4 vs IPv6. In MPLS they are carried on different FECs. Presumably we need to have a different S-Label for each IP type. This implies a different s/n for each IP type, but since it seems unlikely that we need to maintain them in a common sequence that looks to be OK.

## [9.2.](#) Operation at a Relay node (S-PE)

At a relay node packet forwarding follows the same process used in a PW S-PE [[RFC5659](#)], save for the additional processing required for any required PREF processing.

The packet is received from the incoming LSP and any T-labels which are still present are popped. The S-label is looked up in the

platform label table to determine the forwarding parameters. If PREF processing is required the process described in [Section 8](#) is followed, and if required the locally held sequence number information associated with the S-label is updated to avoid future duplicate forwarding.

For each copy of the packet to be forwarded the S-label is swapped for the S-label required by either the next relay node, or by the egress edge node. For each copy of the packet, the T-label(s) needed to reach the next relay node or the egress edge node is pushed and the packet forwarded towards its next hop in the MPLS layer.

## [9.3.](#) Operation at an Egress Edge (PE) Node

At an egress edge node, the S-Label is looked up in the platform label table and is used to determine the egress packet processing parameters. The sequence number in d-CW and the recorded sequence number history is compared to perform any required ER processing [Section 8.2](#). If the packet is not eliminated, the d-CW is stripped.

If the packet is to be treated as an IP packet, this is processed in the normal way for an IP packet egressing an MPLS tunnel (for example the data-link header is constructed) and the packet forwarded towards its destination.

Editor's note: Do we need to look the IP address up in a forwarding table, if so, which one, or is the next hop a forwarding parameter. Is the MAC address a forwarding parameter, or do we need to run ARP as normal?

If the packet is to be treated as an Ethernet packet it is forwarded unmodified, save for the addition of the CRC as described in [[RFC4448](#)].

#### [9.4.](#) Operation at a Transit(P) Node

Operation at a transit (P) node is normal MPLS forwarding. The outer label is either swapped or popped as required, and the packet is forwarded along the LSP. If an entropy label is present in the label stack, this may be used by the Equal Cost Multi-Path (ECMP) selection process. No other label is inspected as part of forwarding.

The Traffic Class field and/or incoming LSP transport label may be used to indicate how the packet is to be processed and queued including which forwarding resources are to be applied to the packet ([RFC3270](#)). Any of the methods of constructing a physical LSP (for RSVP-TE signaling or MPLS-TP style controller based configuration) or a virtual LSP (Segment Routing

[\[I-D.ietf-spring-segment-routing-mpls\]](#)) may be used to indicate not only the next hop, but the priority of processing and any physical resources dedicated to the service [\[I-D.bryant-rtgwg-enhanced-vpn\]](#).

### [10.](#) Other DetNet data plane considerations

#### [10.1.](#) Class of Service

Class and quality of service, i.e., CoS and QoS, are terms that are often used interchangeably and confused. In the context of DetNet, CoS is used to refer to mechanisms that provide traffic forwarding treatment based on aggregate group basis and QoS is used to refer to mechanisms that provide traffic forwarding treatment based on a specific DetNet flow basis. Examples of existing network level CoS mechanisms include DiffServ which is enabled by IP header differentiated services code point (DSCP) field [\[RFC2474\]](#) and MPLS label traffic class field [\[RFC5462\]](#), and at Layer-2, by IEEE 802.1p priority code point (PCP).

CoS for DetNet flows carried in PWs and MPLS is provided using the existing MPLS Differentiated Services (DiffServ) architecture [\[RFC3270\]](#). Both E-LSP and L-LSP MPLS DiffServ modes MAY be used to support DetNet flows. The Traffic Class field (formerly the EXP field) of an MPLS label follows the definition of [\[RFC5462\]](#) and [\[RFC3270\]](#). The Uniform, Pipe, and Short Pipe DiffServ tunneling and TTL processing models are described in [\[RFC3270\]](#) and [\[RFC3443\]](#) and

MAY be used for MPLS LSPs supporting DetNet flows. MPLS ECN MAY also be used as defined in ECN [[RFC5129](#)] and updated by [[RFC5462](#)].

One additional consideration for DetNet nodes which support CoS services is that they MUST ensure that the CoS service classes do not impact the congestion protection and latency control mechanisms used to provide DetNet QoS. This requirement is similar to requirement for MPLS LSRs to that CoS LSPs do not impact the resources allocated to TE LSPs via [[RFC3473](#)].

## [10.2](#). Quality of Service

Quality of Service (QoS) mechanisms for flow specific traffic treatment typically includes a guarantee/agreement for the service, and allocation of resources to support the service. Example QoS mechanisms include discrete resource allocation, admission control, flow identification and isolation, and sometimes path control, traffic protection, shaping, policing and remarking. Example protocols that support QoS control include Resource ReSerVation Protocol (RSVP) [[RFC2205](#)] and RSVP-TE [[RFC3209](#)] and [[RFC3473](#)]. The existing MPLS mechanisms defined to support CoS [[RFC3270](#)] can also be used to reserve resources for specific traffic classes.

In addition to explicit routes [Section 11.2](#), and packet replication and elimination, described in [Section 8](#) above, DetNet provides zero congestion loss and bounded latency and jitter. As described in [[I-D.ietf-detnet-architecture](#)], there are different mechanisms that maybe used separately or in combination to deliver a zero congestion loss service. These mechanisms are provided by the either the MPLS or IP layers, and may be combined with the mechanisms defined by the underlying network layer such as 802.1TSN.

A baseline set of QoS capabilities for DetNet flows carried over MPLS can be provided with Traffic Engineering (MPLS-TE) [[RFC3209](#)] and [[RFC3473](#)]. TE LSPs can also support explicit routes (path pinning). Current service definitions for packet TE LSPs can be found in "Specification of the Controlled Load Quality of Service", [[RFC2211](#)], "Specification of Guaranteed Quality of Service", [[RFC2212](#)], and "Ethernet Traffic Parameters", [[RFC6003](#)]. Additional service definitions are expected in future documents to support the full range of DetNet services. In all cases, the existing label-based marking mechanisms defined for TE-LSPs and even E-LSPs are use to

support the identification of flows requiring DetNet QoS.

Packets that are marked with a DetNet Class of Service value, but that have not been the subject of a completed reservation, can disrupt the QoS offered to properly reserved DetNet flows by using resources allocated to the reserved flows. Therefore, the network nodes of a DetNet network MUST:

- o Defend the DetNet QoS by discarding or remarking (to a non-DetNet CoS) packets received that are not the subject of a completed reservation.
- o Not use a DetNet reserved resource, e.g. a queue or shaper reserved for DetNet flows, for any packet that does not carry a DetNet Class of Service marker.

### [10.3.](#) Bidirectional traffic

Some DetNet applications generate bidirectional traffic. Using MPLS definitions [[RFC5654](#)] there are associated bidirectional flows, and co-routed bidirectional flows. MPLS defines a point-to-point associated bidirectional LSP as consisting of two unidirectional point-to-point LSPs, one from A to B and the other from B to A, which are regarded as providing a single logical bidirectional transport path. This would be analogous of standard IP routing, or PWs running over two reciprocal unidirectional LSPs. MPLS defines a point-to-point co-routed bidirectional LSP as an associated bidirectional LSP which satisfies the additional constraint that its two unidirectional component LSPs follow the same path (in terms of both nodes and

links) in both directions. An important property of co-routed bidirectional LSPs is that their unidirectional component LSPs share fate. In both types of bidirectional LSPs, resource allocations may differ in each direction. The concepts of associated bidirectional flows and co-routed bidirectional flows can be applied to DetNet flows as well. PWs [[RFC3985](#)] are always created as bidirectional constructs.

While the MPLS data planes must support bidirectional DetNet flows, there are no special bidirectional features with respect to the data plane other than need for the two directions take the same paths. Fate sharing and associated vs co-routed bidirectional flows can be

managed at the control level. Note, that there is no stated requirement for bidirectional DetNet flows to be supported using the same MPLS Labels in each direction, and indeed to do so would introduce significant implementation issues. Control mechanisms will be need to support such bidirectional flows DetNet over MPLS, but such mechanisms are out of scope of this document. An example control plane solution for MPLS can be found in [\[RFC7551\]](#).

#### [10.4.](#) Layer 2 addressing and QoS Considerations

Editor's note: Add references needed by this section

The Time-Sensitive Networking (TSN) Task Group of the IEEE 802.1 Working Group have defined (and are defining) a number of amendments to IEEE 802.1Q [\[IEEE8021Q\]](#) that provide zero congestion loss and bounded latency in bridged networks. IEEE 802.1CB [\[IEEE8021CB\]](#) defines packet replication and elimination functions that should prove both compatible with and useful to, DetNet networks.

As is the case for DetNet, a Layer 2 network node such as a bridge may need to identify the specific DetNet flow to which a packet belongs in order to provide the TSN/DetNet QoS for that packet. It also will likely need a CoS marking, such as the priority field of an IEEE Std 802.1Q VLAN tag, to give the packet proper service.

Although the flow identification methods described in IEEE 802.1CB [\[IEEE8021CB\]](#) are flexible, and in fact, include IP 5-tuple identification methods, the baseline TSN standards assume that every Ethernet frame belonging to a TSN stream (i.e. DetNet flow) carries a multicast destination MAC address that is unique to that flow within the bridged network over which it is carried. Furthermore, IEEE 802.1CB [\[IEEE8021CB\]](#) describes three methods by which a packet sequence number can be encoded in an Ethernet frame.

Ensuring that the proper Ethernet VLAN tag priority and destination MAC address are used on a DetNet/TSN packet may require further

clarification of the customary L2/L3 transformations carried out by routers and edge label switches. Edge nodes may also have to move sequence number fields among Layer 2, PW, and IPv6 encapsulations.

#### [11.](#) Management and control considerations



While management plane and control planes are traditionally considered separately, from the Data Plane perspective there is no practical difference based on the origin of flow provisioning information. This document therefore does not distinguish between information provided by a control plane protocol, e.g., RSVP-TE [[RFC3209](#)] and [[RFC3473](#)], or by a network management mechanisms, e.g., RestConf [[RFC8040](#)] and YANG [[RFC7950](#)].

Editor's note: Not sure if RSVP-TE can be used, indeed unsure what routing protocols can be use other than to create point to point MPLS transport paths. Normally we construct a a single path through the network with RSVP-TE, but here we need to construct an explicit mesh at the DetNet layer. The classical routing protocols are not really capable of constructing graphs of the sort needed here either.

#### 11.1. S-Label assignment and distribution

A mechanism based on the existing PW label distribution protocol [[RFC8077](#)] can be used to exchange labels between DetNet nodes. The protocol may however need extending depending on the preferred format of the DetNet flow identifiers.

A mechanism to create the flow graph through the relay nodes will need to be specified. Initially this is likely to be based on a network controller of some sort rather than a distributed routing protocol.

The details of the control plane protocol solution required for the label distribution and the management of the label number space are out of scope of this document.

#### 11.2. Explicit routes

Editor's note describe the applicability of explicit routes as a method of constructing paths

### 12. Security considerations

The security considerations of DetNet in general are discussed in [[I-D.ietf-detnet-architecture](#)] and [[I-D.sdt-detnet-security](#)]. Other security considerations will be added in a future version of this draft.

### 13. IANA considerations

This document makes no IANA requests.

#### 13.1. Acknowledgments

We acknowledge the authors of [draft-ietf-detnet-dp-sol-01](#).

### 14. References

#### 14.1. 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>>.
- [RFC4448] Martini, L., Ed., Rosen, E., El-Aawar, N., and G. Heron, "Encapsulation Methods for Transport of Ethernet over MPLS Networks", [RFC 4448](#), DOI 10.17487/RFC4448, April 2006, <<https://www.rfc-editor.org/info/rfc4448>>.
- [RFC5085] Nadeau, T., Ed. and C. Pignataro, Ed., "Pseudowire Virtual Circuit Connectivity Verification (VCCV): A Control Channel for Pseudowires", [RFC 5085](#), DOI 10.17487/RFC5085, December 2007, <<https://www.rfc-editor.org/info/rfc5085>>.

#### 14.2. Informative References

- [I-D.bryant-rtgwg-enhanced-vpn]  
Bryant, S. and J. Dong, "Enhanced Virtual Private Networks (VPN+)", [draft-bryant-rtgwg-enhanced-vpn-01](#) (work in progress), October 2017.
- [I-D.ietf-detnet-architecture]  
Finn, N., Thubert, P., Varga, B., and J. Farkas, "Deterministic Networking Architecture", [draft-ietf-detnet-architecture-04](#) (work in progress), October 2017.
- [I-D.ietf-detnet-dp-alt]  
Korhonen, J., Farkas, J., Mirsky, G., Thubert, P., Zhuangyan, Z., and L. Berger, "DetNet Data Plane Protocol and Solution Alternatives", [draft-ietf-detnet-dp-alt-00](#) (work in progress), October 2016.

Internet-Draft

Deterministic MPLS

March 2018

[I-D.ietf-detnet-dp-sol]

Korhonen, J., Andersson, L., Jiang, Y., Finn, N., Varga, B., Farkas, J., Bernardos, C., Mizrahi, T., and L. Berger, "DetNet Data Plane Encapsulation", [draft-ietf-detnet-dp-sol-01](#) (work in progress), January 2018.

[I-D.ietf-spring-segment-routing-mpls]

Bashandy, A., Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing with MPLS data plane", [draft-ietf-spring-segment-routing-mpls-12](#) (work in progress), February 2018.

[I-D.sdt-detnet-security]

Mizrahi, T., Grossman, E., Hacker, A., Das, S., Dowdell, J., Austad, H., Stanton, K., and N. Finn, "Deterministic Networking (DetNet) Security Considerations", [draft-sdt-detnet-security-01](#) (work in progress), July 2017.

[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, <<https://www.rfc-editor.org/info/rfc2205>>.

[RFC2211] Wroclawski, J., "Specification of the Controlled-Load Network Element Service", [RFC 2211](#), DOI 10.17487/RFC2211, September 1997, <<https://www.rfc-editor.org/info/rfc2211>>.

[RFC2212] Shenker, S., Partridge, C., and R. Guerin, "Specification of Guaranteed Quality of Service", [RFC 2212](#), DOI 10.17487/RFC2212, September 1997, <<https://www.rfc-editor.org/info/rfc2212>>.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", [RFC 2474](#), DOI 10.17487/RFC2474, December 1998, <<https://www.rfc-editor.org/info/rfc2474>>.

[RFC3032] Rosen, E., Tappan, D., Fedorkow, G., Rekhter, Y., Farinacci, D., Li, T., and A. Conta, "MPLS Label Stack Encoding", [RFC 3032](#), DOI 10.17487/RFC3032, January 2001, <<https://www.rfc-editor.org/info/rfc3032>>.

- [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, <<https://www.rfc-editor.org/info/rfc3209>>.

- [RFC3270] Le Faucheur, F., Wu, L., Davie, B., Davari, S., Vaananen, P., Krishnan, R., Cheval, P., and J. Heinanen, "Multi-Protocol Label Switching (MPLS) Support of Differentiated Services", [RFC 3270](#), DOI 10.17487/RFC3270, May 2002, <<https://www.rfc-editor.org/info/rfc3270>>.
- [RFC3443] Agarwal, P. and B. Akyol, "Time To Live (TTL) Processing in Multi-Protocol Label Switching (MPLS) Networks", [RFC 3443](#), DOI 10.17487/RFC3443, January 2003, <<https://www.rfc-editor.org/info/rfc3443>>.
- [RFC3473] Berger, L., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Signaling Resource ReserVation Protocol-Traffic Engineering (RSVP-TE) Extensions", [RFC 3473](#), DOI 10.17487/RFC3473, January 2003, <<https://www.rfc-editor.org/info/rfc3473>>.
- [RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", [RFC 3985](#), DOI 10.17487/RFC3985, March 2005, <<https://www.rfc-editor.org/info/rfc3985>>.
- [RFC4206] Kompella, K. and Y. Rekhter, "Label Switched Paths (LSP) Hierarchy with Generalized Multi-Protocol Label Switching (GMPLS) Traffic Engineering (TE)", [RFC 4206](#), DOI 10.17487/RFC4206, October 2005, <<https://www.rfc-editor.org/info/rfc4206>>.
- [RFC4385] Bryant, S., Swallow, G., Martini, L., and D. McPherson, "Pseudowire Emulation Edge-to-Edge (PWE3) Control Word for Use over an MPLS PSN", [RFC 4385](#), DOI 10.17487/RFC4385, February 2006, <<https://www.rfc-editor.org/info/rfc4385>>.
- [RFC5129] Davie, B., Briscoe, B., and J. Tay, "Explicit Congestion Marking in MPLS", [RFC 5129](#), DOI 10.17487/RFC5129, January

2008, <<https://www.rfc-editor.org/info/rfc5129>>.

- [RFC5462] Andersson, L. and R. Asati, "Multiprotocol Label Switching (MPLS) Label Stack Entry: "EXP" Field Renamed to "Traffic Class" Field", [RFC 5462](#), DOI 10.17487/RFC5462, February 2009, <<https://www.rfc-editor.org/info/rfc5462>>.
- [RFC5654] Niven-Jenkins, B., Ed., Brungard, D., Ed., Betts, M., Ed., Sprecher, N., and S. Ueno, "Requirements of an MPLS Transport Profile", [RFC 5654](#), DOI 10.17487/RFC5654, September 2009, <<https://www.rfc-editor.org/info/rfc5654>>.

Bryant & Chen

Expires September 6, 2018

[Page 25]

---

Internet-Draft

Deterministic MPLS

March 2018

- [RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", [RFC 5659](#), DOI 10.17487/RFC5659, October 2009, <<https://www.rfc-editor.org/info/rfc5659>>.
- [RFC5921] Bocci, M., Ed., Bryant, S., Ed., Frost, D., Ed., Levrau, L., and L. Berger, "A Framework for MPLS in Transport Networks", [RFC 5921](#), DOI 10.17487/RFC5921, July 2010, <<https://www.rfc-editor.org/info/rfc5921>>.
- [RFC6003] Papadimitriou, D., "Ethernet Traffic Parameters", [RFC 6003](#), DOI 10.17487/RFC6003, October 2010, <<https://www.rfc-editor.org/info/rfc6003>>.
- [RFC7551] Zhang, F., Ed., Jing, R., and R. Gandhi, Ed., "RSVP-TE Extensions for Associated Bidirectional Label Switched Paths (LSPs)", [RFC 7551](#), DOI 10.17487/RFC7551, May 2015, <<https://www.rfc-editor.org/info/rfc7551>>.
- [RFC7950] Bjorklund, M., Ed., "The YANG 1.1 Data Modeling Language", [RFC 7950](#), DOI 10.17487/RFC7950, August 2016, <<https://www.rfc-editor.org/info/rfc7950>>.
- [RFC8040] Bierman, A., Bjorklund, M., and K. Watsen, "RESTCONF Protocol", [RFC 8040](#), DOI 10.17487/RFC8040, January 2017, <<https://www.rfc-editor.org/info/rfc8040>>.
- [RFC8077] Martini, L., Ed. and G. Heron, Ed., "Pseudowire Setup and

Maintenance Using the Label Distribution Protocol (LDP)",  
STD 84, [RFC 8077](https://www.rfc-editor.org/info/rfc8077), DOI 10.17487/RFC8077, February 2017,  
<<https://www.rfc-editor.org/info/rfc8077>>.

#### Authors' Addresses

Stewart Bryant  
Huawei Technologies

Email: [stewart.bryant@gmail.com](mailto:stewart.bryant@gmail.com)

Mach Chen  
Huawei Technologies

Email: [mach.chen@huawei.com](mailto:mach.chen@huawei.com)