DetNet
Internet-Draft
Intended status: Informational
Expires: January 9, 2017

J. Korhonen, Ed.
Broadcom
J. Farkas
G. Mirsky
Ericsson
P. Thubert
Cisco
Y. Zhuang
Huawei
L. Berger
LabN

July 8, 2016

# DetNet Data Plane Protocol and Solution Alternatives
draft-dt-detnet-dp-alt-01

## Abstract

This document identifies existing IP and MPLS, and other encapsulations that run over IP and/or MPLS data plane technologies that can be considered as the base line solution for deterministic networking data plane definition.

## 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 January 9, 2017.

## Copyright Notice

Table of Contents

## 1. Introduction

Deterministic Networking (DetNet) [I-D.ietf-detnet-problem-statement] provides a capability to carry unicast or multicast data flows for real-time applications with extremely low data loss rates, timely delivery and bounded packet delay variation [I-D.finn-detnet-architecture]. The deterministic networking Quality of Service (QoS) is expressed as 1) the minimum and the maximum end-to-end latency from sender (talker) to receiver (listener), and 2) probability of loss of a packet. Only the worst-case values for the mentioned parameters are concerned.

There are three techniques to achieve the QoS required by deterministic networks:

- Bandwidth reservation and enforcement,
- explicit routes,
- packet loss protection, initially provided by replication and elimination.

This document identifies existing IP and Multiprotocol Label Switching (MPLS) [RFC3031], layer-2 or layer-3 encapsulations and transport protocols that could be considered as foundations for a deterministic networking data plane. The full scope of the deterministic networking data plane solution is considered including, as appropriate: quality of service (QoS); Operations, Administration and Maintenance (OAM); and time synchronization among other criteria described in Section 4.

This document does not select a deterministic networking data plane protocol. It does, however, elaborate what it would require to adapt and use a specific protocol as the deterministic networking data plane solution. This document is only concerned with data plane considerations and, specifically, with topics that potentially impact potential deterministic networking aware data plane hardware. Control plane considerations are out of scope of this document.

## 2. Terminology

This document will eventually use the terminology established in the DetNet architecture [I-D.finn-detnet-architecture]. Currently the following terms are used:

DetNet Reliability

: A set of mechanisms to increase the probability of lossless (i.e., zero loss) DetNet flow delivery across a network. Example mechanisms include packet replication and duplicate elimination.

Transit Node

A node that provides link layer and network layer switching across multiple links and/or sub-networks. Transit nodes provide packet forwarding services to DetNet nodes. An MPLS LSR, or IP router are example transit nodes.

Relay Node

A DetNet Service aware middle box that interconnects different network layer protocols or networks (instances). A relay node also understands enough of the DetNet service and service parameter semantics to make an intelligent processing (e.g., forwarding) decision. It may provide service supporting functions such as DetNet reliability.

Edge Node

A relay node with application level knowledge (i.e., basically a "proxy" node). Egde nodes include DetNet application level functions and are needed when interfacing (or inter-working) with nodes and end systems that are not DetNet-enabled.

## 3. DetNet Data Plane Overview

A "Deterministic Network" will be composed of DetNet enabled nodes i.e., End Systems, Edge Nodes, Relay Nodes and collectively deliver DetNet services. DetNet enabled nodes are interconnected via Transit Nodes (i.e., routers) which support DetNet, but are not DetNet service aware. Transit nodes see DetNet nodes as end points. All DetNet enabled nodes are connect to sub-networks, where a point-to-point link is also considered as a simple sub-network. These sub-networks will provide DetNet compatible service for support of DetNet traffic. Examples of sub-networks include IEEE 802.1TSN and OTN. Of course, multi-layer DetNet systems may also be possible, where one DetNet appears as a sub-network, and provides service to, a higher layer DetNet system. A simple DetNet concept network is shown in Figure 1.

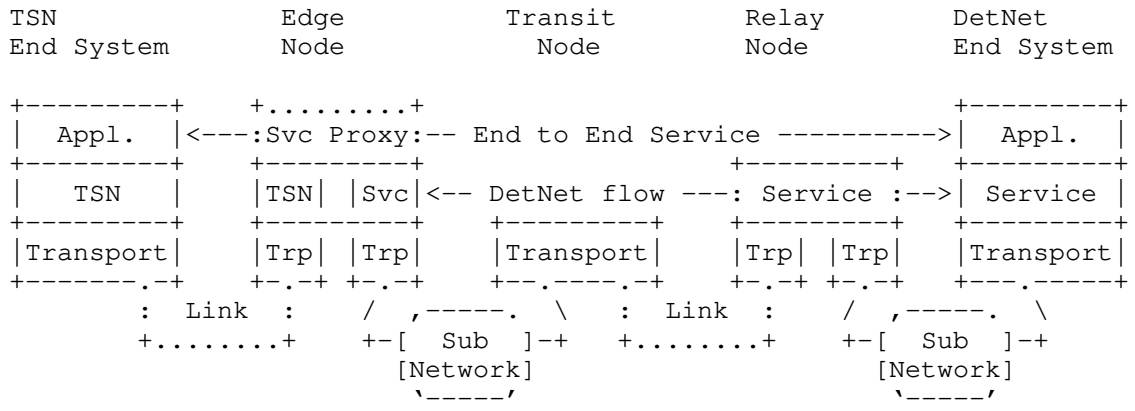



Figure 1: A Simple DetNet Enabled Network

The protocol stack model of the data plane described in [I-D.finn-detnet-architecture] defines functional primitives for ingress and egress packets, which are used by the three techniques (see Section 1) to ensure deterministic forwarding. [I-D.finn-detnet-architecture] does not specify the relationship between the DetNet Service and Transport layers used in this document to investigate data plane options as explained in the following. The goal of this document is to evaluate possible data plane technologies and compare their characteristics from DetNet perspective.

The DetNet data plane is logically divided into two layers (also see Figure 2):

DetNet Service Layer

The DetNet service layer provides adaptation of DetNet services. It is composed of a shim layer to carry deterministic flow specific attributes, which are needed during forwarding. DetNet enabled end systems originate and terminate the DetNet Service layer and are peers at the DetNet Service layer. DetNet relay and edge nodes also implement DetNet Service layer functions. The DetNet service layer is used to deliver traffic end to end across a DetNet domain.

DetNet Transport Layer

The DetNet transport layer is required on all DetNet nodes. All DetNet nodes are end points and the transport layer. Non-DetNet service aware transit nodes deliver traffic between DetNet nodes. The DetNet transport layer operates below and supports the DetNet Service layer.

Distinguishing the function of these two DetNet data plane layers helps to explore and evaluate various combinations of the data plane solutions available. This separation of DetNet layers, while helpful, should not be considered as formal requirement. For example, some technologies may violate these strict layers and still be able to deliver a DetNet service.

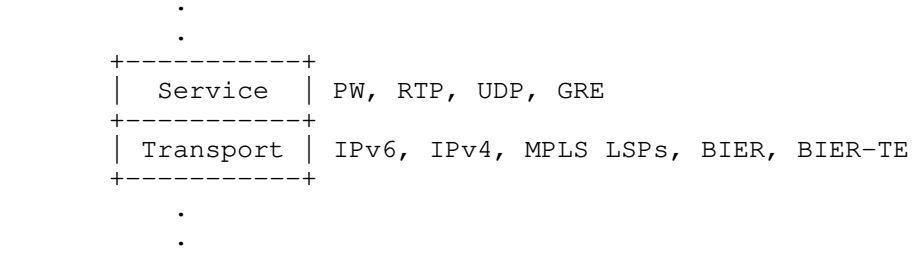



Figure 2: DetNet adaptation to data plane

The two logical layers defined here aim to help to identify which data plane technology can be used for what purposes in the DetNet context. This layering is similar to the data plane concept of MPLS, where some part of the label stack is "Service" specific (e.g., PW labels, VPN labels) and an other part is "Transport" specific (e.g, LSP label, TE label(s)).

In some networking scenarios, the end system initially provides a DetNet flow encapsulation, which contains all information needed by DetNet nodes (e.g., RTP based DetNet flow transported over a native UDP/IP network). In other scenarios, the encapsulation formats might differ significantly. As an example, a CPRI "application's" IQ data mapped directly to Ethernet frames may have to be transported over an MPLS based packet switched network (PSN).

There are many valid options to create a data plane solution for DetNet traffic by selecting a technology approach for the DetNet Service layer and also selecting a technology approach for the DetNet Transport layer. There are a high number of valid combinations. Therefore, not the combinations but the different technologies are evaluated along the criteria collected in Section 4. Different criteria apply for the DetNet Service layer and the DetNet Transport layer, however, some of the criteria are valid for both layers.

One of the most fundamental differences between different potential data plane options is the basic addressing and headers used by DetNet end systems. For example, is the basic service a Layer 2 (e.g., Ethernet) or Layer 3 (i.e., IP) service. This decision impacts how

DetNet end systems are addressed, and the basic forwarding logic for the DetNet Service layer

## 3.1. Example DetNet Service Scenarios

In an attempt to illustrate a DetNet date plane, this document uses the Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3) [RFC5254] reference model shown in Figure 3 as the foundation for different DetNet data plane deployment options and how layering could work. Other reference models are possible but not covered in this document. Note that other technologies can be also used to implement DetNet, Multi-Segment PW is only used here to illustrate functions, features and layering from the perspective of the architecture.

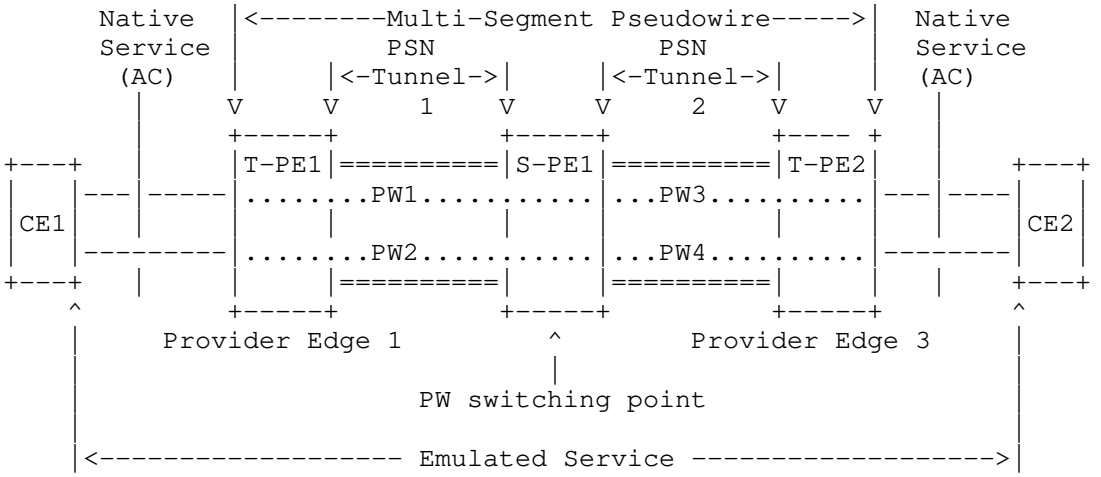



Figure 3: Pseudo Wire switching reference model

Figure 4 illustrates how DetNet can provide services for IEEE 802.1TSN end systems over a DetNet enabled network. The edge nodes insert and remove required DetNet data plane encapsulation. The 'X' in the edge and relay nodes represents a potential DetNet flow packet replication and elimination point. This conceptually parallels L2VPN services, and could leverage existing related solutions as discussed below.

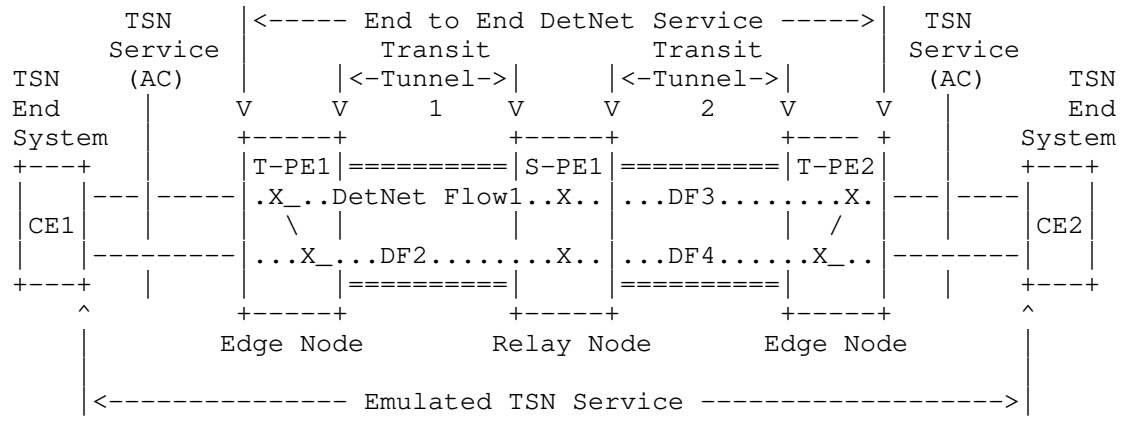



Figure 4: IEEE 802.1TSN over DetNet

Figure 5 illustrates how end to end native DetNet service can be provided. In this case, the end systems are able to send and receive native DetNet flows. For example, as PseudoWire (PW) encapsulated IP. Like earlier 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 replacing IP with MPLS or tunneling IP over MPLS (e.g., via L3VPNs), or simply interconnect network domains.

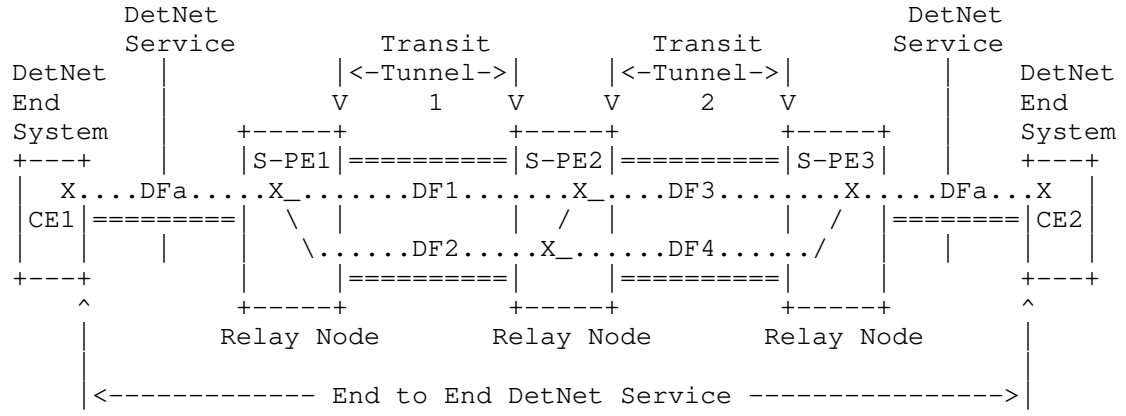



Figure 5: Native DetNet

Figure 6 illustrates how a IEEE 802.1TSN end system could communicate with a native DetNet end system through an edge node which provides a TSN to DetNet inter-working capability. The edge node would add and remove required DetNet data plane encapsulation as well as provide any needed address mapping. As in previous figures, the 'X' in the

end systems, edge and relay nodes represents potential DetNet flow packet duplication and elimination points.

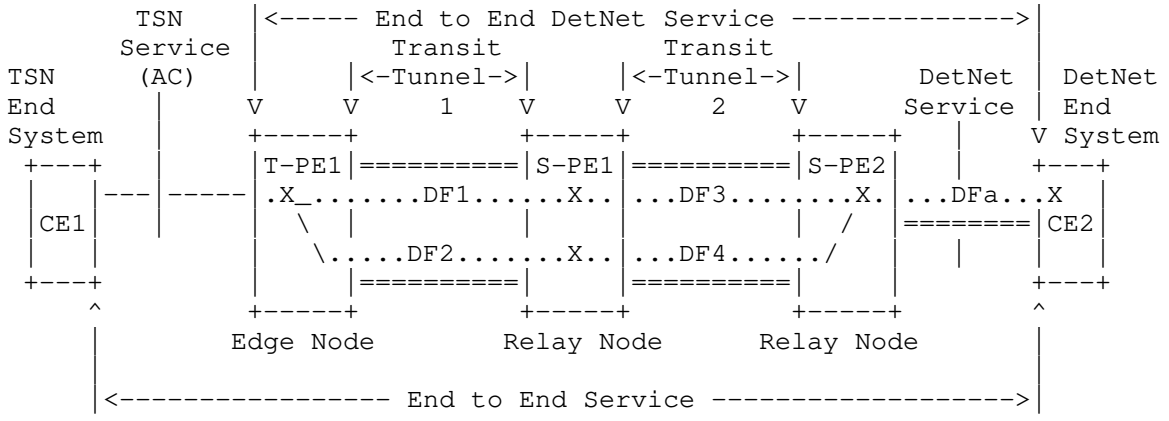



Figure 6: IEEE 802.1TSN to native DetNet

## 4. Criteria for data plane solution alternatives

This section provides criteria to help to evaluate potential options. Each deterministic networking data plane solution alternative is described and evaluated using the criteria described in this section. The used criteria enumerated in this section are selected so that they highlight the existence or lack of features that are expected or seen important to a solution alternative for the data plane solution.

The criteria for the DetNet Service layer:

#1 Encapsulation and overhead
#2 Flow identification (Service ID part of the DetNet flows)
#3 Packet sequencing (sequence number)
#5 Packet replication and elimination (note: only the packet deletion for DetNet Reliability)
#6 Operations, Administration and Maintenance (capabilities)
#8 Class and quality of service capabilities (DetNet Service specific)
#10 Technical maturity

The criteria for the DetNet Transport layer:

#1 Encapsulation and overhead
#2 Flow identification
#4 Explicit routes (network path)
#5 Packet replication and elimination (note: only the packet replication and/or flow merging for DetNet Reliability)

#6 Operations, Administration and Maintenance (capabilities, performance management, packet traceability)
#8 Class and quality of service capabilities (DetNet Transport specific)
#9 Packet traceability (can be part of OAM)
#10 Technical maturity

[Editor's Note: numbering is off because #7 is removed.]

[Editor's Note: #9 should(?) be integrated into #6.]

Most of the criteria is relevant for both the DetNet Service and DetNet Transport layers. However, different aspects of the same criteria may relevant for different layers, for example, as it is the case with criteria #5 Packet replication and elimination.

### 4.1. #1 Encapsulation and overhead

Encapsulation and overhead is related to how the DetNet data plane carries DetNet flow. In several cases a DetNet flow has to be encapsulated inside other protocols, for example, when transporting a layer-2 Ethernet frame over an IP transport network. In some cases a tunneling like encapsulation can be avoided by underlying transport protocol translation, for example, translating layer-2 Ethernet frame including addressing and flow identification into native IP traffic. Last it is possible that talkers and listeners handle deterministic flows natively in layer-3. This criteria concerns what is the encapsulation method the solution alternative support: tunneling like encapsulation, protocol translation or native layer-3 transport. In addition to the encapsulation mechanism this criteria is also concerned of the processing and specifically the encapsulate header overhead.

### 4.2. #2 Flow identification

The solution alternative has to provide means to identify specific deterministic flows. The flow identification can, for example, be explicit field in the data plane encapsulation header or implicitly encoded into the addressing scheme of the used data plane protocol or their combination. This criteria concerns the availability and details of deterministic flow identification the data plane protocol alternative has.

### 4.3. #3 Packet sequencing

[Editor's note: is in order delivery a strict requirement? if so, it should be stated as such and separately from any other requirement. There are multiple ways to solve this criteria.]

The solution alternative has to provide means for end systems to number packets sequentially and transport that sequencing information along with the sent packets. In addition to possible reordering packets other important uses for sequencing are detecting duplicates and lost packets.

In a case of intentional packet duplication a combination of flow identification and packet sequencing allows for detecting and discarding duplicates at the receiver (see Section 4.5 for more details). This criteria concerns the availability and details of the packet sequencing capabilities the data plane protocol alternative has.

### 4.4. #4 Explicit routes

The solution alternative has to provide a mechanism(s) for establishing explicit routes that all packets belonging to a deterministic flow will follow. The explicit route can be seen as a form of source routing or a pre-reserved path e.g., using some network management procedure. It should be noted that the explicit route does not need to be detailed to a level where every possible intermediate node along the path is part of the named explicit route. RSVP-TE [RFC3209] supports explicit routes, and typically provides pinned data paths for established LSPs. At Layer-2, the IEEE 802.1Qca [IEEE802.1Qca] specification defines how to do explicit path control in a bridged network and its IETF counter part is defined in [RFC7813]. This criteria concerns the available mechanisms for explicit routes for the data plane protocol alternative.

### 4.5. #5 Packet replication and elimination

Packet replication and elimination is the first method being considered to provide DetNet reliability. The objective for supporting packet replication and elimination is to enable hitless (or lossless) 1+1 protection, which is also called Seamless Redundancy in [I-D.finn-detnet-architecture]. Data plane solutions need to meet this objective independent of the used solution. In other words, a packet replication and elimination is one identified method for a data plane solution to provide seamless redundancy (e.g., DetNet Reliability). Other methods, if so identified, are also permissible.

The solution alternative has to provide means for end systems and/or relay systems to be able to replicate packets, and later eliminate all but one of the replicas, at multiple points in the network in order to ensure that one (or more) equipment failure event(s) still leave at least one path intact for a deterministic networking flow. The goal is to enable hitless 1+1 protection in a way that no packet

gets lost or there is no ramp up time when either one of the paths fails for one reason or another.

Another concern regarding packet replication is how to enforce replicated packets to take different route or path while the final destination still remains the same. With strict source routing, all the intermediate hops are listed and paths can be guaranteed to be non-overlapping. Loose source routing only signals some of the intermediate hops and it takes additional knowledge to ensure that there is no single point of failure.

The IEEE 802.1CB (seamless redundancy) [IEEE8021CB] is an example of Ethernet-based solution that defines packet sequence numbering, packet replication, and duplicate packet identification and deletion. The deterministic networking data plane solution alternative at layer-3 has to provide equivalent functionality. This criteria concerns the available mechanisms for packet replication and duplicate deletion the data plane protocol alternative has.

### 4.6. #6 Operations, Administration and Maintenance

The solution alternative should demonstrate an availability of appropriate standardized OAM tools that can be extended for deterministic networking purposes with a reasonable effort, when required. The OAM tools do not necessarily need to be specific to the data plane protocol as it could be the case, for example, with MPLS-based data planes. But any OAM-related implications or requirements on data plane hardware must be considered.

The OAM includes but is not limited to tools listed in the requirements for overlay networks [I-D.ooamdt-rtgwg-ooam-requirement]. Specifically, the performance management requirements are of interest at both service and transport layers.

### 4.7. #8 Class and quality of service capabilities

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 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).

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] (RSVP) and RSVP-TE [RFC3209] and [RFC3473].

A critical DetNet service enabled by QoS (and perhaps CoS) is delivering zero congestion loss. There are different mechanisms that maybe used separately or in combination to deliver a zero congestion loss service. The key aspect of this objective is that DetNet packets are not discarded due to congestion at any point in a DetNet aware network.

In the context of the data plane solution there should be means for flow identification, which then can be used to map a flow against specific resources and treatment in a node enforcing the QoS. Hereto, certain aspects of CoS and QoS may be provided by the underlying sub-net technology, e.g., actual queuing or IEEE 802.3x priority flow control (PFC).

### 4.8. #9 Packet traceability

For the network management and specifically for tracing implementation or network configuration errors any means to find out whether a packet is a replica, which node performed replication, and which path was intended for the replica, can be very useful. This criteria concerns the availability of solutions for tracing packets in the context of data plane protocol alternative. Packet traceability can also be part of OAM.

### 4.9. #10 Technical maturity

The technical maturity of the data plane solution alternative is crucial, since it basically defines the effort, time line and risks involved for the use of the solution in deployments. For example, the maturity level can be categorized as available immediately, available with small extensions, available with re-purposing/ redefining portions of the protocol or its header fields. Yet another important measure for maturity is the deployment experience. This criteria concerns the maturity of the data plane protocol alternative as the solution alternative. This criteria is particularly important given, as previously noted, that the DetNet data plane solution is expected to impact, i.e., be supported in, hardware.

# 5. Data plane solution alternatives

The following sections describe and rate deterministic data plane solution alternatives. In "Analysis and Discussion" section each alternative is evaluated against the criteria given in Section 4 and rated using the following: (M)eets the criteria, (W)ork needed, and (N)ot suitable or too much work envisioned.

## 5.1. DetNet Transport layer technologies

### 5.1.1. Native IPv6 transport

#### 5.1.1.1. Solution description

This section investigates the application of native IPv6 [RFC2460] as the data plane for deterministic networking along the criteria collected in Section 4.

The application of higher OSI layer headers, i.e., headers deeper in the packet, can be considered. Two aspects have to be taken into account for such solutions. (i) Those header fields can be encrypted. (ii) Those header fields are deeper in the packet, therefore, routers have to apply deep packet inspection. See further details in Section 5.2.5.

#### 5.1.1.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

IPv6 can encapsulate DetNet Service layer headers (and associated DetNet flow payload) like any other upper-layer header indicated by the Next Header. The fixed header of an IPv6 packet is 40 bytes [RFC2460]. This overhead is bigger if any Extension Header is used, and a generic behaviour for host and forwarding nodes is specified in [RFC7045]. However, the exact overhead (Section 4.1) depends on what solution is actually used to provide DetNet features, e.g., explicit routing or DetNet Reliability if any of these is applied.

IPv6 has two types of Extension Headers that are processed by intermediate routers between the source and the final destination and may be of interest for the data plane signaling, the Routing Header that is used to direct the traffic via intermediate routers in a strict or loose source routing way, and the Hop-by-Hop Options Header that carries optional information that must be examined by every node along a packet's delivery path. The Hop-by-Hop Options Header, when present, must immediately follow the

IPv6 Header and it is not possible to limit its processing to the end points of Source Routed segments.

IPv6 also provides a Destination Options Header that is used to carry optional information to be examined only by a packet's destination node(s). The encoding of the options used in the Hop-by-Hop and in the Destination Options Header indicates the expected behavior when a processing IPv6 node does not recognize the Option Type, e.g. skip or drop; it should be noted that due to performance restrictions nodes may ignore the Hop-by-Hop Option Header, drop packets containing a Hop-by-Hop Option Header, or assign packets containing a Hop-by-Hop Option Header to a slow processing path [I-D.ietf-6man-rfc2460bis] (e.g. punt packets from hardware to software forwarding which is highly detrimental to the performance).

The creation of new Extension Headers that would need to be processed by intermediate nodes is strongly discouraged. In particular, new Extension Header(s) having hop-by-hop behavior must not be created or specified. New options for the existing Hop-by-Hop Header should not be created or specified unless no alternative solution is feasible [RFC6564].

#2 Flow identification (W)

The 20-bit flow label field of the fixed IPv6 header is suitable to distinguish different deterministic flows. But guidance on the use of the flow label provided by [RFC6437] places restrictions on how the flow label can be used. In particular, labels should be chosen from an approximation to a discrete uniform distribution. Additionally, existing implementations generally do not open APIs to control the flow label from the upper layers.

Alternatively, the Flow identification could be transported in a new option in the Hop-by-Hop Options Header.

#4 Explicit routes (W)

One possibility is for a Software-Defined Networking (SDN) [RFC7426] based approach to be applied to compute, establish and manage the explicit routes, leveraging Traffic Engineering (TE) extensions to routing protocols [RFC5305] [RFC7752] and evolving to the Path Computation Element (PCE) Architecture [RFC5440], though a number of issues remain to be solved [RFC7399].

Segment Routing (SR) [I-D.ietf-spring-segment-routing] is a new initiative to equip IPv6 with explicit routing capabilities. The

idea for the DetNet data plane would be to apply SR to IPv6 with the addition of a new type of routing extension header [I-D.ietf-6man-segment-routing-header] to explicitly signal the path in the data plane between the source and the destination, and/or between replication points and elimination points if this functionality is used.

#5 Packet replication and elimination (W)

The functionality of replicating a packet exists in IPv6 but is limited to multicast flows. In order to enforce replicated packets to take different routes, IP-in-IP encapsulation and Segment Routing could be leveraged to signal a segment in a packet. A replication point would insert a different routing header in each copy it makes, the routing header providing explicitly the hops to the elimination point for that particular replica of the packet, in a strict or in a loose source routing fashion. An elimination point would pop the routing headers from the various copies it gets and forward or receive the packet if it is the final destination.

#6 Operations, Administration and Maintenance (M/W)

IPv6 enjoys the existing toolbox for generic IP network management. However, IPv6 specific management features are still not at the level comparable to that of IPv4. Particular areas of concerns are those that are IPv6 specific, for example, related to neighbor discovery protocol (ND), stateless address autoconfiguration (SLAAC), subscriber identification, and security. While the standards are already mostly in place the implementations in deployed equipment can be lacking or inadequate for commercial deployments. This is larger issue with older existing equipment.

#8 Class and quality of service capabilities (W)

IPv6 provides support for CoS and QoS. CoS is provided by DiffServ which is enabled by IP header differentiated services code point (DSCP) and QoS is defined as part of RSVP [RFC2205]. DiffServ support is widely available, while RSVP for IP packets is generally not supported.

#9 Packet traceability (W)

The traceability of replicated packets involves the capability to resolve which replication point issued a particular copy of a

packet, which segment was intended for that replica, and which particular packet of which particular flow this is. Sequence also depends on the sequencing mechanism. As an example, the replication point may be indicated as the source of the packet if IP-in-IP encapsulation is used to forward along segments. Another alternate to IP-in-IP tunneling along segments would be to protect the original source address in a destination option similar to the Home Address option [RFC6275] and then use the address of the replication point as source in the IP header.

The traceability also involves the capability to determine if a particular segment is operational. While IPv6 as such has no support for reversing a path, it appears source route extensions such as the one defined for segment routing could be used for tracing purposes. Though it is not a usual practice, IPv6 [RFC2460] expects that a Source Route path may be reversed, and the standard insists that a node must not include the reverse of a Routing Header in the response unless the received Routing Header was authenticated.

#10 Technical maturity (M/W)

IPv6 has been around about 20 years. However, large scale global and commercial IPv6 deployments are rather new dating only few years back to around 2012. While IPv6 has proven itself for best effort traffic, DiffServ usage is less common and QoS capabilities are not currently present. Additional, there are number of small issues to work on as they show up once operations experience grows.

The Cisco 6Lab site [1] provides information on IPv6 deployment per country, indicating figures for prefixes, transit AS, content and users. Per this site, many countries, including Canada, Brazil, the USA, Germany, France, Japan, Portugal, Sweden, Finland, Norway, Greece, and Ecuador, achieve a deployment ratio above 30 percent, and the overall adoption reported by Google Statistics [2] is now above 10 percent.

#### 5.1.1.3. Summary

IPv6 supports a significant portion of the identified DetNet data plane criteria today. There are aspects of the DetNet data plane that are not fully supported, notably QoS, but these can be incrementally added or supplemented by the underlying sub-network layer. IPv6 may be a choice as the DetNet Transport layer in networks where other technologies such as MPLS are not deployed.

### 5.1.2. Native IPv4 transport

#### 5.1.2.1. Solution description

IPv4 [RFC0791] is in principle the same as IPv6, except that it has a smaller address space. However, IPv6 was designed around the fact that extension headers are an integral part of the protocol and operation from the beginning, although the practice may some times prove differently [RFC7872]. IPv4 does support header options, but these have historically not been supported on in hardware-based forwarding so are generally blocked or handled at a much slower rate. In either case, the use of IP header options is generally avoided. In the context of deterministic networking data plane solutions the major difference between IPv4 and IPv6 seems to be the practical support for header extensibility. Anything below and above the IP header independent of the version is practically the same.

#### 5.1.2.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

The fixed header of an IPv4 packet is 20 bytes [RFC0791]. IP options add overhead, but are not generally used and are not considered as part of this document.

#2 Flow identification (W)

The IPv4 header has a 16-bit identification field that was originally intended for assisting fragmentation and reassembly of IPv4 packets as described in [RFC0791]. The identification field has also been proposed to be used for actually identifying flows between two IP addresses and a given protocol for detecting and removing duplicate packets [RFC1122]. However, recent update [RFC6864] to both [RFC0791] and [RFC1122] restricts the use of IPv4 identification field only to fragmentation purposes.

The IPv4 also has a stream identifier option [RFC0791], which contains a 16-bit SATNET stream identifier. However, the option has been deprecated [RFC6814]. The conclusion is that stream identification does not work nicely with IPv4 header alone and a traditional 5-tuple identification might not also be enough in a case of a flow duplication or encrypted flows. For a working solution, upper layer protocol headers such as RTP or PWs may be required for unambiguous flow identification. There is also emerging work within the IETF that may provide new flow identification alternatives.

#4 Explicit routes (W)

IPv4 has two source routing option specified: the loose source and record route option (LSRR), and the strict source and record route option (SSRR) [RFC0791]. The support of these options in the Internet is questionable but within a closed network the support may be assumed. But as both these options use IP header options, which are generally not supported in hardware, use of these options are questionable. Of course, the same options of SDN and SR approaches discussed above for IPv6 may be equally applicable to IPv4.

#5 Packet replication and elimination (W/N)

The functionality of replicating a packet exists in IPv4 but is limited to multicast flows. In general the issue regarding the IPv6 packet replication also applies to IPv4. Duplicate packet detection for IPv4 is studied in [RFC6621] to a great detail in the context of simplified multicast forwarding. In general there is no good way to detect duplicated packets for IPv4 without additional upper layer protocol support.

#6 Operations, Administration and Maintenance (M)

IPv4 enjoys the extensive and "complete" existing toolbox for generic IP network management.

#8 Class and quality of service capabilities (M/W)

IPv4 provides support for CoS and QoS. CoS is provided by DiffServ which is enabled by IP header differentiated services code point (DSCP) and QoS is defined as part of RSVP [RFC2205]. DiffServ support is widely available, while RSVP for IP packets is generally not supported.

#9 Packet traceability (W)

The IPv4 has similar needs and requirements for traceability as IPv6 (see Section 5.1.1.2). The IPv4 has a traceroute option [RFC6814] that could be used to record the route the packet took. However, the option has been deprecated [RFC6814].

#10 Technical maturity (M/W)

IPv4 can be considered mature technology with over 30 years of implementation, deployment and operations experience. As with IPv6, today’s commercial implementations and deployments of IPv4 generally lack any support for QoS.

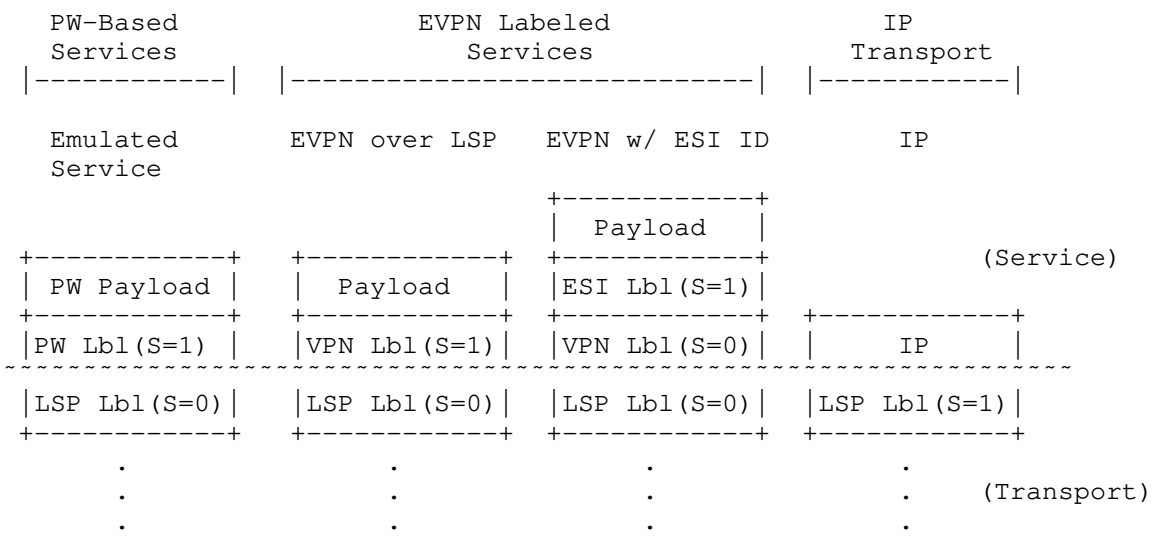
#### 5.1.2.3. Summary

The IPv4 has specifications to support most of the identified DetNet data plane criteria today. However, several of those have already been deprecated or their wide support is not guaranteed. The DetNet data plane criteria that are not fully supported could be incrementally added or supplemented by the underlying sub-network layer. Unfortunately, the IPv4 has had limited success getting its extensions deployed at large. However, introducing new extensions might have a better success in closed networks (like DetNet) than in Internet. Due to the popularity of the IPv4, it should be considered as a potential choice for the DetNet Transport layer.

### 5.1.3. Multiprotocol Label Switching (MPLS)

Multiprotocol Label Switching Architecture (MPLS) [RFC3031] and its variants, MPLS with Traffic Engineering (MPLS-TE) [RFC3209] and [RFC3473], and MPLS Transport Profile (MPLS-TP) [RFC5921] is a widely deployed technology that switches traffic based on MPLS label stacks [RFC3032] and [RFC5960]. MPLS is the foundation for Pseudowire-based services Section 5.2.3 and emerging technologies such as Bit-Indexed Explicit Replication (BIER) Section 5.1.4 and Source Packet Routing [3].

MPLS supports the equivalent of both the DetNet Service and DetNet Transport layers, and provides a very rich set of mechanisms that can be reused directly, and perhaps augmented in certain cases, to deliver DetNet services. At the DetNet Transport layer, MPLS provides forwarding, protection and OAM services. At the DetNet Service Layer it provides client service adaption, directly, via Pseudowires Section 5.2.3 and via other label-like mechanisms such as EPVN Section 5.2.4. A representation of these options are shown in Figure 7.




~~~~~~~~~~~ denotes DetNet Service <-> DetNet Transport layer boundary

Figure 7: MPLS-based Services

MPLS can be controlled in a number of ways including via a control plane, via the management plane, or via centralized controller (SDN) based approaches. MPLS also provides standard control plane reference points. Additional information on MPLS architecture and control can be found in [RFC5921]. A summary of MPLS control plane related functions can be found in [RFC6373]. The remainder of this section will focus [RFC6373]. The remainder of this section will focus on the MPLS transport data plane, additional information on the MPLS service data plane can be found below in Section 5.2.2.

5.1.3.1. Solution description

The following draws heavily from [RFC5960].

Encapsulation and forwarding of packets traversing MPLS LSPs follows standard MPLS packet encapsulation and forwarding as defined in [RFC3031], [RFC3032], [RFC5331], and [RFC5332].

Data plane Quality of Service capabilities are included in the MPLS in the form of Traffic Engineered (TE) LSPs [RFC3209] and the MPLS Differentiated Services (DiffServ) architecture [RFC3270]. Both E-LSP and L-LSP MPLS DiffServ modes are defined. The Traffic Class field (formerly the EXP field) of an MPLS label follows the definition of [RFC5462] and [RFC3270].
~~~~~~~~~~~

Except for transient packet reordering that may occur, for example, during fault conditions, packets are delivered in order on L-LSPs, and on E-LSPs within a specific ordered aggregate.

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.

Equal-Cost Multi-Path (ECMP) load-balancing is possible with MPLS LSPs and can be avoided using a number of techniques. The same holds for Penultimate Hop Popping (PHP).

MPLS includes the following LSP types:

- Point-to-point unidirectional
- Point-to-point associated bidirectional
- Point-to-point co-routed bidirectional
- Point-to-multipoint unidirectional

Point-to-point unidirectional LSPs are supported by the basic MPLS architecture [RFC3031].

A point-to-point associated bidirectional LSP between LSRs A and B consists of two unidirectional point-to-point LSPs, one from A to B and the other from B to A, which are regarded as a pair providing a single logical bidirectional transport path.

A point-to-point co-routed bidirectional LSP is a point-to-point associated bidirectional LSP with the additional constraint that its two unidirectional component LSPs in each direction follow the same path (in terms of both nodes and links). An important property of co-routed bidirectional LSPs is that their unidirectional component LSPs share fate.

A point-to-multipoint unidirectional LSP functions in the same manner in the data plane, with respect to basic label processing and packet-switching operations, as a point-to-point unidirectional LSP, with one difference: an LSR may have more than one (egress interface, outgoing label) pair associated with the LSP, and any packet it transmits on the LSP is transmitted out all associated egress interfaces. Point-to-multipoint LSPs are described in [RFC4875] and [RFC5332]. TTL processing and exception handling for point-to-multipoint LSPs is the same as for point-to-point LSPs.

Additional data plane capabilities include Linear Protection, [RFC6378] and [RFC7271]. And the in progress work on MPLS support for time synchronization [I-D.ietf-mpls-residence-time].

5.1.3.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

There are two perspectives to consider when looking at encapsulation. The first is encapsulation to support services. These considerations are part of the DetNet service layer and are covered below, see Sections 5.2.3 and 5.2.4.

The second perspective relates to encapsulation, if any, is needed to transport packets across network. In this case, the MPLS label stack, [RFC3032] is used to identify flows across a network. MPLS labels are compact and highly flexible. They can be stacked to support client adaptation, protection, network layering, source routing, etc.

The number of DetNet Transport layer specific labels is flexible and support a wide range of applicable functions and MPLS domain characteristics (e.g., TE-tunnels, Hierarchical-LSPs, etc.).

#2 Flow identification (M)

MPLS label stacks provide highly flexible ways to identify flows. Basically, they enable the complete separation of traffic classification from traffic treatment and thereby enable arbitrary combinations of both.

For the DetNet flow identification the MPLS label stack can be used to support n-layers of DetNet flow identification. For example, using dedicated LSP per DetNet flow would simplify flow identification for intermediate transport nodes, and additional hierarchical LSPs could be used to facilitate scaling.

#4 Explicit routes (M)

MPLS supports explicit routes based on how LSPs are established, e.g., via TE explicit routes [RFC3209]. Additional, but not required, capabilities are being defined as part of Segment Routing (SR) [I-D.ietf-spring-segment-routing].

#5 Packet replication and elimination (M/W)

MPLS as DetNet Transport layer supports the replication via point-to-multipoint LSPs. Duplicate elimination is not provided and would need to be provided within a Detnet function. However, at

the MPLS LSP level, there are mechanisms defined to provide 1+1 protection. The current definitions [RFC6378] and [RFC7271] use OAM mechanisms to support and coordinate protection switching and packet loss is possible during a switch. While such this level of protection may be sufficient for many DetNet applications, when truly hitless (i.e., zero loss) switching is required, additional mechanisms will be needed. It is expected that these additional mechanisms will be defined at a DetNet layer.

#6 Operations, Administration and Maintenance (M)

MPLS already includes a rich set of OAM functions at both the Service and Transport Layers. This includes LSP ping [ref] and those enabled via the MPLS Generic Associated Channel [RFC5586] and registered by IANA [4].

#8 Class and quality of service capabilities (M/W)

As previously mentioned, Data plane Quality of Service capabilities are included in the MPLS in the form of Traffic Engineered (TE) LSPs [RFC3209] and the MPLS Differentiated Services (DiffServ) architecture [RFC3270]. Both E-LSP and L-LSP MPLS DiffServ modes are defined. The Traffic Class field (formerly the EXP field) of an MPLS label follows the definition of [RFC5462] and [RFC3270]. One potential open area of work is synchronized, time based scheduling. Another is shaping, which is generally not supported in shipping MPLS hardware.

#9 Packet traceability (M)

MPLS supports multiple tracing mechanisms. A control based one is defined in [RFC3209]. An OAM based mechanism is defined in MPLS On-Demand Connectivity Verification and Route Tracing [RFC6426].

#10 Technical maturity (M)

MPLS as a mature technology that has been widely deployed in many networks for many years. Numerous vendor products and multiple generations of MPLS hardware have been built and deployed.

#### 5.1.3.3. Summary

MPLS is a mature technology that has been widely deployed. Numerous vendor products and multiple generations of MPLS hardware have been built and deployed. MPLS LSPs support a significant portion of the identified DetNet data plane criteria today. Aspects of the DetNet data plane that are not fully supported can be incrementally added. It's worth noting that a number of limitations are in shipping hardware, versus at the protocol specification level, e.g., shaping.

### 5.1.4. Bit Indexed Explicit Replication (BIER)

Bit Indexed Explicit Replication [I-D.ietf-bier-architecture] (BIER) is a network plane replication technique that was initially intended as a new method for multicast distribution. In a nutshell, a BIER header includes a bitmap that explicitly signals the listeners that are intended for a particular packet, which means that 1) the sender is aware of the individual listeners and 2) the BIER control plane is a simple extension of the unicast routing as opposed to a dedicated multicast data plane, which represents a considerable reduction in OPEX. For this reason, the technology faces a lot of traction from Service Providers. Section 5.1.4 discusses the applicability of BIER for replication in the DetNet.

The simplicity of the BIER technology makes it very versatile as a network plane signaling protocol. Already, a new Traffic Engineering variation is emerging that uses bits to signal segments along a TE path. While the more classical BIER is mainly a multicast technology that typically leverages a unicast distributed control plane through IGP extensions, BIER-TE is mainly a unicast technology that leverages a central computation to setup path, compute segments and install the mapping in the intermediate nodes. Section 5.1.5 discusses the applicability of BIER-TE for replication, traceability and OAM operations in DetNet.

Bit-Indexed Explicit Replication (BIER) layer may be considered to be included into Deterministic Networking data plane solution. Encapsulation of a BIER packet in MPLS network presented in Figure 8

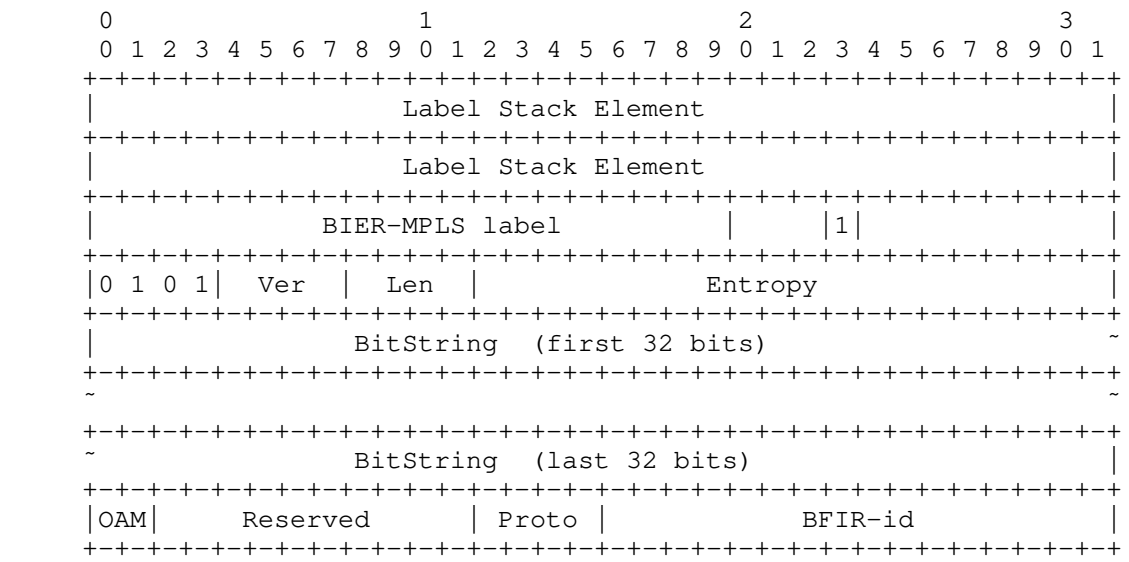



Figure 8: BIER packet in MPLS encapsulation

5.1.4.1. Solution description

The DetNet may be presented in BIER as distinctive payload type with its own Proto(col) ID. Then it is likely that DetNet will have the header that would identify:

- o Version;
- o Sequence Number;
- o Timestamp;
- o Payload type, e.g. data vs. OAM.

DetNet node, collocated with BFIR, may use multiple BIER sub-domains to create replicated flows. Downstream DetNet nodes, collocated with BFER, would terminate redundant flows based on Sequence Number and/or Timestamp information. Such DetNet may be BFER in one BIER sub-domain and BFIR in another. Thus DetNet flow would traverse several BIER sub-domains.

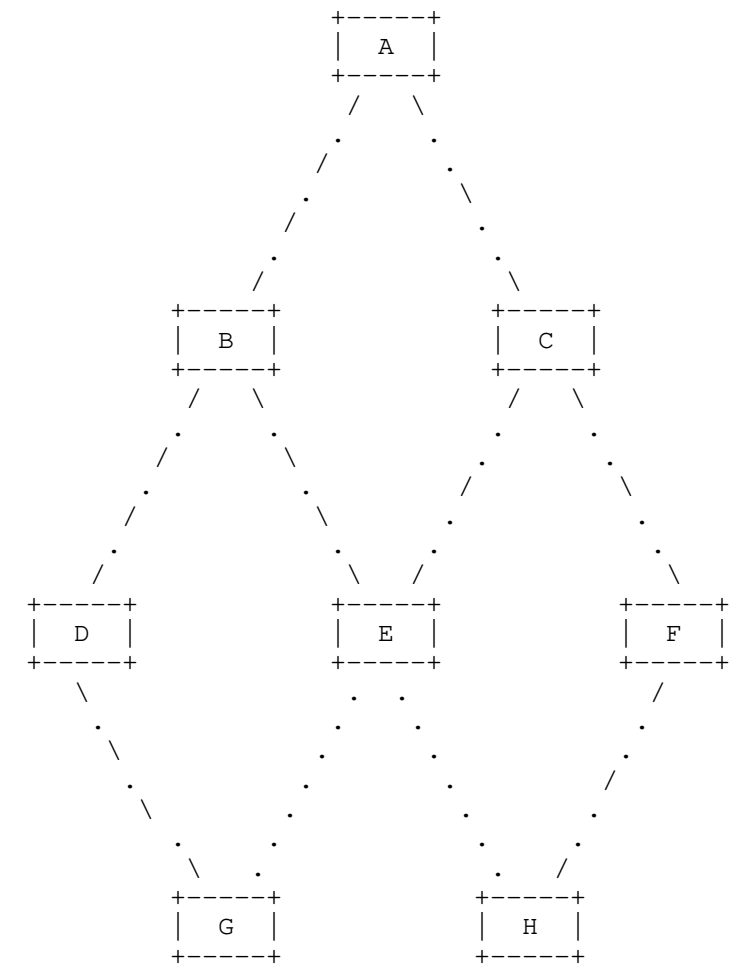



Figure 9: DetNet in BIER domain

Consider DetNet flow that must traverse BIER enabled domain from A to G and H. DetNet may use three BIER subdomains:

- A-B-D-E-G (dash-dot): A is BFIR, E and G are BFERs,
- A-C-E-F-H (dash-double-dot): A is BFIR, E and H are BFERs,
- E-G-H (dotted): E is BFIR, G and H are BFERs.

DetNet node A sends DetNet into red and purple BIER sub-domains. DetNet node E receives DetNet packet and sends into green sub-domain while terminating duplicates and those that deemed too-late.

DetNet nodes G and H receive DetNet flows, terminate duplicates and those that are too-late.

5.1.4.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

BIER over MPLS network encapsulation (will refer as "BIER over MPLS" further for short), Figure 8, is being defined [I-D. ietf-bier-mpls-encapsulation] within the BIER working group.

#2 Flow identification (M)

Flow identification and separation can be achieved through use of BIER domains and/or Entropy value in the BIER over MPLS, Figure 8.

#4 Explicit routes (M)

Explicit routes may be used as underlay for BIER domain. BIER underlay may be calculated using PCE and instantiated using any southbound mechanism.

#5 Packet replication and elimination (M/W)

Packet replication, as indicated by its name, is core function of the Bit-Indexed Explicit Replication. Elimination of the duplicates and/or too-late packets cannot be done within BIER sub-domain but may be done at DetNet overlay at the edge of the BIER sub-domain.

#6 Operations, Administration and Maintenance (M/W)

BIER over MPLS guarantees that OAM is fate-sharing, i.e. in-band with a data flow being monitored or measured. Additionally, BIER over MPLS enables passive performance measurement, e.g. with the marking method [I-D.mirsky-bier-pmmm-oam]. Some OAM protocols, e.g. can be applied and used in BIER over MPLS as demonstrated [I-D.ooamdt-rtgwg-oam-gap-analysis], while new protocols being worked on, e.g. ping/traceroute [I-D.kumarzheng-bier-ping] or Path MTU Discovery [I-D.mirsky-bier-path-mtu-discovery].

#8 Class and quality of service capabilities (M/W)

Class of Service can be inherited from the underlay of the particular BIER sub-domain. Quality of Service, i.e. scheduling and bandwidth reservations can be used among other constrains in calculating explicit path for the BIER sub-domain’s underlay.

#9 Packet traceability (W)

Ability to do passive performance measurement by using OAM field of the BIER over MPLS, Figure 8, is unmatched and significantly simplifies truly passive tracing of selected flows and packets within them.

#10 Technical maturity (W)

The BIER over MPLS is nearing finalization within the BIER WG and several experimental implementations are expected soon.

5.1.4.3. Summary

BIER over MPLS supports a significant portion of the identified DetNet data plane requirements, including controlled packet replication, traffic engineering, while some requirements, e.g. duplicate and too-late packet elimination may be realized as function of the DetNet overlay. BIER over MPLS is a viable candidate as the DetNet Transport layer in MPLS networks.

5.1.5. BIER - Traffic Engineering (BIER-TE)

An alternate use of Bit-Indexed Explicit Replication (BIER) uses bits in the BitString to represent adjacencies as opposed to destinations, as discussed in BIER Traffic Engineering (TE) [I-D.eckert-bier-te-arch].

The proposed function of BIER-TE in the DetNet data plane is to control the process of replication and elimination, as opposed to the identification of the flows or and the sequencing of packets within a flow.

At the path ingress, BIER-TE identifies the adjacencies that are activated for this packet (under the rule of the controller). At the egress, BIER-TE is used to identify the adjacencies where transmission failed. This information is passed to the controller, which in turn can modify the active adjacencies for the next packets.

The value is that the replication can be controlled and monitored in a loop that may involve an external controller, with the granularity of a packet and an adjacency .

5.1.5.1. Solution description

BIER-TE enables to activate the replication and elimination functions in a manner that is abstract to the data plane forwarding information. An adjacency, which is represented by a bit in the BIER

header, can correspond in the data plane to an Ethernet hop, a Label Switched Path, or it can correspond to an IPv6 loose or strict source routed path.

In a nutshell, BIER-TE is used as follows:

- A controller computes a complex path, sometimes called a track, which takes the general form of a ladder. The steps and the side rails between them are the adjacencies that can be activated on demand on a per-packet basis using bits in the BIER header.

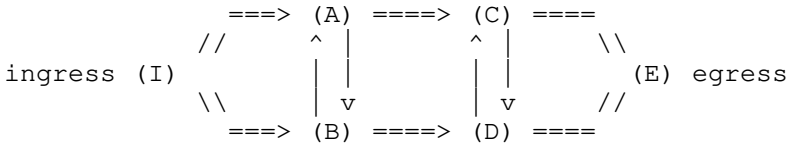



Figure 10: Ladder Shape with replication and elimination Points

- The controller assigns a BIER domain, and inside that domain, assigns bits to the adjacencies. The controller assigns each bit to a replication node that sends towards the adjacency, for instance the ingress router into a segment that will insert a routing header in the packet. A single bit may be used for a step in the ladder, indicating the other end of the step in both directions.

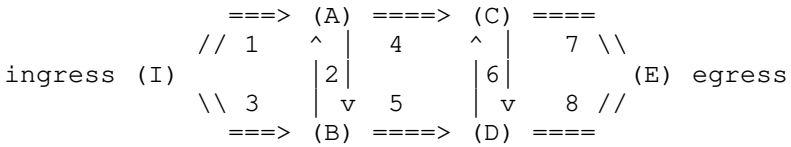



Figure 11: Assigning Bits

- The controller activates the replication by deciding the setting of the bits associated with the adjacencies. This decision can be modified at any time, but takes the latency of a controller round trip to effectively take place. Below is an example that uses replication and elimination to protect the A->C adjacency.

| Bit # | Adjacency | Owner | Example Bit Setting |
|---|---|---|---|
| 1 | I->A | I | 1 |
| 2 | A->B | A | 1 |
| | B->A | B | |
| 3 | I->C | I | 0 |
| 4 | A->C | A | 1 |
| 5 | B->D | B | 1 |
| 6 | C->D | C | 1 |
| | D->C | D | |
| 7 | C->E | C | 1 |
| 8 | D->E | D | 0 |

replication and elimination Protecting A->C

Table 1: Controlling Replication

- The BIER header with the controlling BitString is injected in the packet by the ingress node of the deterministic path. That node may act as a replication point, in which case it may issue multiple copies of the packet

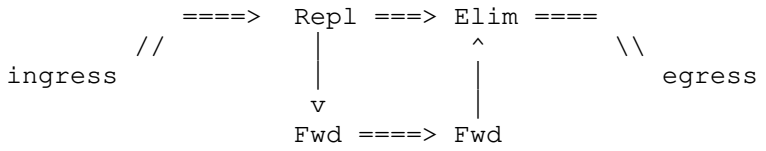



Figure 12: Enabled Adjacencies

- For each of its bits that is set in the BIER header, the owner replication point resets the bit and transmits towards the associated adjacency; to achieve this, the replication point copies the packet and inserts the relevant data plane information, such as a source route header, towards the adjacency that corresponds to the bit

| Adjacency | BIER BitString |
|---|---|
| I->A | 01011110 |
| A->B | 00011110 |
| B->D | 00010110 |
| D->C | 00010010 |
| A->C | 01001110 |

BitString in BIER Header as Packet Progresses

Table 2: BIER-TE in Action

- Adversely, an elimination node on the way strips the data plane information and performs a bitwise AND on the BitStrings from the various copies of the packet that it has received, before it forwards the packet with the resulting BitString.

| Operation | BIER BitString |
|---|---|
| D->C | 00010010 |
| A->C | 01001110 |
| | -------- |
| AND in C | 00000010 |
| | |
| C->E | 00000000 |

BitString Processing at Elimination Point C

Table 3: BIER-TE in Action (cont.)

- In this example, all the transmissions succeeded and the BitString at arrival has all the bits reset - note that the egress may be an Elimination Point in which case this is evaluated after this node has performed its AND operation on the received BitStrings).

| Failing Adjacency | Egress BIER BitString |
|---|---|
| I->A | Frame Lost |
| I->B | Not Tried |
| A->C | 00010000 |
| A->B | 01001100 |
| B->D | 01001100 |
| D->C | 01001100 |
| C->E | Frame Lost |
| D->E | Not Tried |

BitString indicating failures

Table 4: BIER-TE in Action (cont.)

- But if a transmission failed along the way, one (or more) bit is never cleared. Table 4 provides the possible outcomes of a transmission. If the frame is lost, then it is probably due to a failure in either I->A or C->E, and the controller should enable I->B and D->E to find out. A BitString of 00010000 indicates unequivocally a transmission error on the A->C adjacency, and a BitString of 01001100 indicates a loss in either A->B, B->D or D->C; enabling D->E on the next packets may provide more information to sort things out.

In more details:

The BIER header is of variable size, and a DetNet network of a limited size can use a model with 64 bits if 64 adjacencies are enough, whereas a larger deployment may be able to signal up to 256 adjacencies for use in very complex paths. Figure 8 illustrates a BIER header as encapsulated within MPLS. The format of this header is common to BIER and BIER-TE.

For the DetNet data plane, a replication point is an ingress point for more than one adjacency, and an elimination point is an egress point for more than one adjacency.

A pre-populated state in a replication node indicates which bits are served by this node and to which adjacency each of these bits corresponds. With DetNet, the state is typically installed by a controller entity such as a PCE. The way the adjacency is signaled in the packet is fully abstracted in the bit representation and must be provisioned to the replication nodes and maintained as a local state, together with the timing or shaping information for the associated flow.

The DetNet data plane uses BIER-TE to control which adjacencies are used for a given packet. This is signaled from the path ingress, which sets the appropriate bits in the BIER BitString to indicate which replication must happen.

The replication point clears the bit associated to the adjacency where the replica is placed, and the elimination points perform a logical AND of the BitStrings of the copies that it gets before forwarding.

As is apparent in the examples above, clearing the bits enables to trace a packet to the replication points that made any particular copy. BIER-TE also enables to detect the failing adjacencies or sequences of adjacencies along a path and to activate additional replications to counter balance the failures.

Finally, using the same BIER-TE bit for both directions of the steps of the ladder enables to avoid replication in both directions along the crossing adjacencies. At the time of sending along the step of the ladder, the bit may have been already reset by performing the AND operation with the copy from the other side, in which case the transmission is not needed and does not occur (since the control bit is now off).

#### 5.1.5.2. Analysis and Discussion

#1 Encapsulation and overhead (W/M)

The size of the BIER header depends on the number of segments in the particular path. It is very concise considering the amount of information that is carried (control of replication, traceability, and measurement of the reliability of the segments).

#2 Flow identification (N)

Some fields in the BIER header could be used to identify the flows but they are not the primary purpose, so it’s probably not a good idea.

#4 Explicit routes (N)

A separate procedure must be used to set up the paths and allocate the bits for the adjacencies. The bits should be distributed as a form of tag by the route setup protocol. This procedure requires more work and is separate from the data plane method that is described here.

#5 Packet replication and elimination (M/W)

The bitmap expresses in a very concise fashion which replication and elimination should take place for a given packet . It also enables to control that process on a per packet basis, depending on the loss that it enables to measure. The net result is that a complex path may be installed with all the possibilities and that the decision of which possibilities are used is controlled in the data plane.

#6 Operations, Administration and Maintenance (W)

The setting of the bits at arrival enables to determine which adjacencies worked and which did not, enabling a dynamic control of the replication and elimination process. This is a form of OAM that is in-band with the data stream as opposed to leveraging separate packets, which is a more accurate information on the reliability of the link for the user.

#8 Class and quality of service capabilities (N)

BIER-TE does not signal that explicitly.

#9 Packet traceability (W)

This is a strong point of the solution. The solution enables to determine which is the current segment that a given packet is expected to traverse, which node performed the replication and which should perform the elimination if any

#10 Technical maturity (W)

Some components of the technology are more mature, e.g. segment routing and BIER. Yet, the overall solution has never been deployed as is not fully defined.

5.1.5.3. Summary

BIER-TE occupies a particular position in the DetNet data plane. In the one hand it is optional, and only useful if replication and elimination is taking place. In the other hand, it has unique capabilities to:

- o control which replication take place on a per packet basis, so that replication points can be configured but not actually utilized
- o trace the replication activity and determine which node replicated a particular packet
- o measure the quality of transmission of the actual data packet along the replication segments and use that in a control loop to adapt the setting of the bits and maintain the reliability.

## 5.2. DetNet Service layer technologies

### 5.2.1. Generic Routing Encapsulation (GRE)

#### 5.2.1.1. Solution description

Generic Routing Encapsulation (GRE) [RFC2784] provides an encapsulation of an arbitrary network layer protocol over another arbitrary network layer protocol. The encapsulation of a GRE packet can be found in Figure 13.

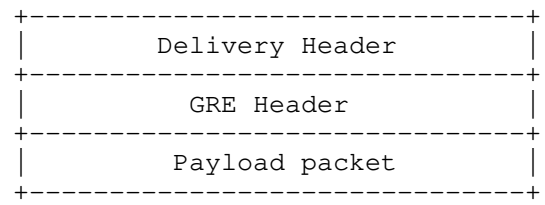



Figure 13: Encapsulation of a GRE packet

Based on RFC2784, [RFC2890] further includes sequencing number and Key in optional fields of the GRE header, which may help to transport DetNet traffic flows over IP networks. The format of a GRE header is presented in Figure 14.

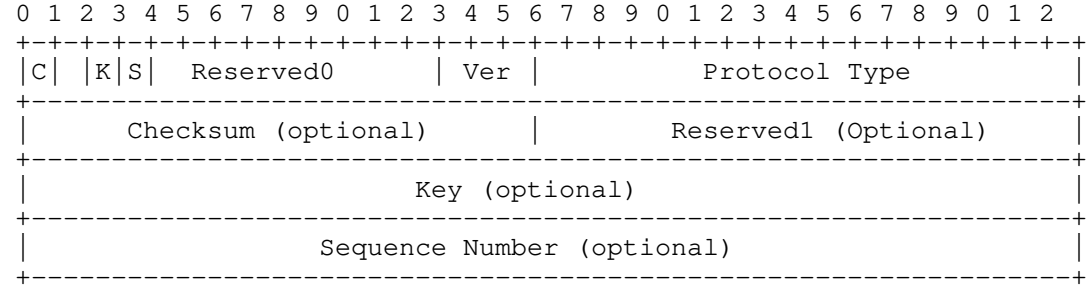



Figure 14: Format of a GRE header

5.2.1.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

GRE can provide encapsulation at the service layer over the transport layer. A new protocol type for DetNet traffic should be allocated as an "Ether Type" in [RFC3232] and in IANA Ethernet Numbers [5]. The fixed header of a GRE packet is 4 octets while the maximum header is 16 octets with optional fields in Figure 14.

#2 Flow identification (W)

There is no flow identification field in GRE header. However, it can rely on the flow identification mechanism applied in the delivery protocols, such as flow identification stated in IP Sections 5.1.1 and 5.1.2 when the delivery protocols are IPv6 and IPv4 respectively. Alternatively, the Key field can also be extended to carry the flow identification. The size of Key field is 4 octets.

#3 Packet sequencing (M)

As stated in Section 5.2.1, GRE provides an optional sequencing number in its header to provide sequencing services for packets. The size of the sequencing number is 32 bits.

#5 Packet replication and elimination (W/N)

GRE has no packet replication and elimination in its header. It can use the transport IPv4/IPv6 protocols at the transport layer to replicate the packets and take the different routes as discussed in Section 5.1.1 and Section 5.1.2. Besides, the GRE header can be extended to indicate the duplicated packets by defining a flag in reserved fields or using the sequencing number of a flow.

#6 Operations, Administration and Maintenance (M)

GRE uses the network management provided by the IP protocols as transport layer.

#8 Class and quality of service capabilities (W)

For the class of service capability, an optional code point field to indicate CoS of the traffic could be added into the GRE header. Otherwise, GRE can reuse the class and quality of service of delivery protocols at transport layer such as IPv6 and IPv4 stated in Section 5.1.1 and Section 5.1.2.

#10 Technical maturity (M)

GRE has been developed over 20 years. The delivery protocol mostly used is IPv4, while the IPv6 support for GRE is to be standardized now in IETF as [RFC7676]. Due to its good extensibility, GRE has also been extended to support network virtualization in Data Center, which is NVGRE [RFC7637].

5.2.1.3. Summary

As a tunneling protocol, GRE can encapsulate a wide variety of network layer protocols over another network layer, which can naturally serve as the service layer protocol for DetNet. Currently, it supports a portion of the Detnet service layer criteria, and still some are not fully supported but can be incrementally added or supported by delivery protocols at as the transport layer. In general, GRE can be a choice as the DetNet service layer and can work with IPv6 and IPv4 as the DetNet Transport layer.

5.2.2. MPLS-based Services for DetNet

MPLS based technologies supports both the DetNet Service and DetNet Transport layers. This, as well as a general overview of MPLS, is covered above in Section 5.1.3. These sections focus on the DetNet Service Layer it provides client service adaption, via Pseudowires Section 5.2.3 and via native and other label-like mechanisms such as EPVN in Section 5.2.4. A representation of these options was previously discussed and is shown in Figure 7.

The following text is adapted from [RFC5921]:

The MPLS native service adaptation functions interface the client layer network service to MPLS. For Pseudowires, these adaptation functions are the payload encapsulation described in Section 4.4 of [RFC3985] and Section 6 of [RFC5659]. For network layer client services, the adaptation function uses the MPLS encapsulation format as defined in [RFC3032].

The purpose of this encapsulation is to abstract the data plane of the client layer network from the MPLS data plane, thus contributing to the independent operation of the MPLS network.

MPLS may itself be a client of an underlying server layer. MPLS can thus also bounded by a set of adaptation functions to this server layer network, which may itself be MPLS. These adaptation functions provide encapsulation of the MPLS frames and for the transparent transport of those frames over the server layer network.

While MPLS service can provided on and true end-system to end-system basis, it's more likely that DetNet service will be provided over Pseudowires as described in Section 5.2.3 or via an EPVN-based service described in Section 5.2.4 .

MPLS labels in the label stack may be used to identify transport paths, see Section 5.1.3, or as service identifiers. Typically a single label is used for service identification.

Packet sequencing mechanisms are added in client-related adaptation processing, see Sections 5.2.3 and 5.2.4.

The MPLS client inherits its Quality of Service (QoS) from the MPLS transport layer, which in turn inherits its QoS from the server (sub-network) layer. The server layer therefore needs to provide the necessary QoS to ensure that the MPLS client QoS commitments can be satisfied.

### 5.2.3. Pseudo Wire Emulation Edge-to-Edge (PWE3)

#### 5.2.3.1. Solution description

Pseudo Wire Emulation Edge-to-Edge (PWE3) [RFC3985] or simply PseudoWires (PW) provide means of emulating the essential attributes and behaviour of a telecommunications service over a packet switched network (PSN) using IP or MPLS transport. In addition to traditional telecommunications services such as T1 line or Frame Relay, PWs also provide transport for Ethernet service [RFC4448] and for generic packet service [RFC6658]. Figure 15 illustrate the reference PWE3 stack model.

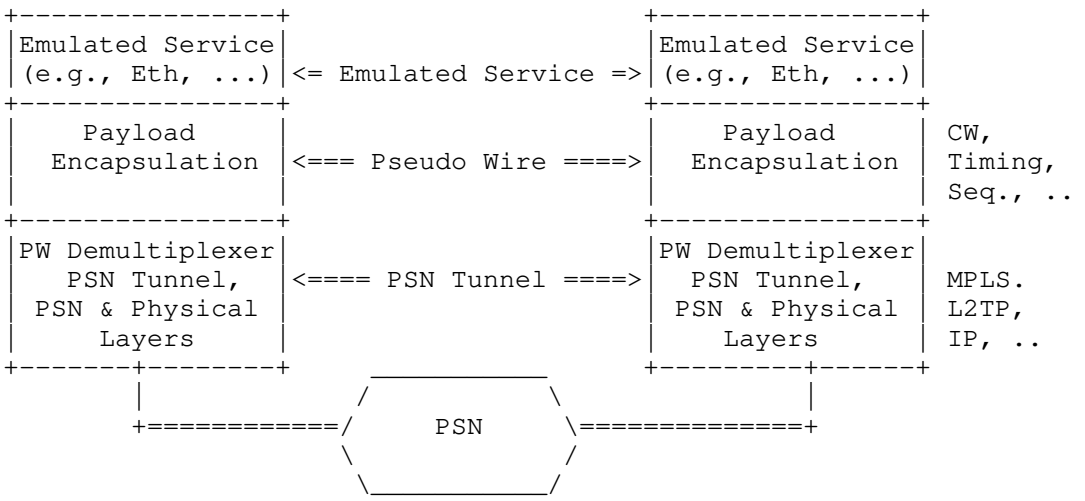



Figure 15: PWE3 protocol stack reference model

PWs appear as a good data plane solution alternative for a number of reasons. PWs are a proven and deployed technology with a rich OAM control plane [RFC4447], and enjoy the toolbox developed for MPLS networks. Furthermore, PWs may have an optional Control Word (CW) as part of the payload encapsulation between the PSN and the emulated service that is, for example, capable of frame sequencing and duplicate detection. The encapsulation layer may also provide timing [RFC5087]. Furthermore, advances DetNet node functions are conceptually already supported by PW framework (with some added functional required), such as the DetNet Relay node modeled after the Multi-Segment PWE3 [RFC5254].

PWs can be also used if the PSN is IP, which enables the application of PWs in networks that do not have MPLS enabled in their core routers. One approach to provide PWs over IP is to provide MPLS over IP in some way and then leverage what is available for PWs over MPLS. The following standard solutions are available both for IPv4 and IPv6 to follow this approach. The different solutions have different overhead as discussed in the following subsection. The MPLS-in-IP encapsulation is specified by [RFC4023]. The IPv4 Protocol Number field or the IPv6 Next Header field is set to 137, which indicates an MPLS unicast packet. (The use of the MPLS-in-IP encapsulation for MPLS multicast packets is not supported.) The MPLS-in-GRE encapsulation is specified in [RFC4023], where the IP header (either IPv4 or IPv6) is followed by a GRE header, which is followed by an MPLS label stack. The protocol type field in the GRE header is set to MPLS Unicast (0x8847) or Multicast (0x8848). MPLS over L2TPv3

over IP encapsulation is specified by [RFC4817]. The MPLS-in-UDP encapsulation is specified by [RFC7510], where the UDP Destination Port indicates tunneled MPLS packet and the UDP Source Port is an entropy value that is generated by the encapsulator to uniquely identify a flow. MPLS-in-UDP encapsulation can be applied to enable UDP-based ECMP (Equal-Cost Multipath) or Link Aggregation. All these solutions can be secured with IPSec.

5.2.3.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

PWs offer encapsulation services practically for any types of payloads over any PSN. New PW types need a code point allocation [RFC4446] and in some cases an emulated service specific document.

Specifically in the case of the MPLS PSN the PW encapsulation overhead is minimal. Typically minimum two labels and a CW is needed, which totals to 12 octets. PW type specific handling might, however, allow optimizations on the emulated service in the provider edge (PE) device's native service processing (NSP) / forwarder function. These optimizations could be used, for example, to reduce header overhead. Ethernet PWs already have rather low overhead [RFC4448]. Without a CW and VLAN tags the Ethernet header gets reduced to 14 octets (minimum Ethernet header overhead is 26).

The overhead is somewhat bigger in case of IP PSN if an MPLS over IP solution is applied to provide PWs. IP adds at least 20 (IPv4) or 40 (IPv6) bytes overhead to the PW over MPLS overhead; furthermore, the GRE, L2TPv3, or UDP header has to be taken into account if any of these further encapsulations is used.

#2 Flow identification (M)

[Editor's note: this criteria has not been checked against the latest view of flow identification after the separation of transport and service layers.]

PWs provide multiple layers of flow identification, especially in the case of the MPLS PSN. The PWs are typically prepended with an endpoint specific PW label that can be used to identify a specific PW per endpoint. Furthermore, the MPLS PSN also uses one or more labels to transport packets over a specific label switched paths (that then would carry PWs). So, a DetNet flow can be identified in this example by the service and transport layer labels. IP (and other) PSNs may need other mechanisms, such as, UDP port

numbers, upper layer protocol header (like RTP) or some IP extension header to provide required flow identification.

#3 Packet sequencing (M)

As mentioned earlier PWs may contain an optional CW that is able to provide sequencing services. The size of the sequence number in the generic CW is 16 bits, which might be, depending on the used link and DetNet flow speed be too little.

#5 Packet replication and elimination (W)

The PW duplicate detection mechanism is already conceptually specified [RFC3985] but no emulated service makes use of it currently.

#6 Operations, Administration and Maintenance (M/W)

PWs have rich control plane for OAM and in a case of the MPLS PSN enjoy the full control plane toolbox developed for MPLS network OAM likewise IP PSN have the full toolbox of IP network OAM tools. There could be, however, need for deterministic networking specific extensions for the mentioned control planes.

#8 Class and quality of service capabilities (M/W)

In a case of IP PSN the 6-bit differentiated services code point (DSCP) field can be used for indicating the class of service [RFC2474] and 2-bit field reserved for the explicit congestion notification (ECN) [RFC3168]. Similarly, in a case of MPLS PSN, there are 3-bit traffic class field (TC) [RFC5462] in the label reserved for for both Explicitly TC-encoded-PSC LSPs (E-LSP) [RFC3270] and ECN [RFC5129]. Due to the limited number of bits in the TC field, their use for QoS and ECN functions restricted and intended to be flexible. Although the QoS/CoS mechanism is already in place some clarifications may be required in the context of deterministic networking flows, for example, if some specific mapping between bit fields have to be done.

When PWs are used over MPLS, MPLS LSPs can be used to provide both CoS (E-LSPs and L-LSPs) and QoS (dedicated TE LSPS).

#10 Technical maturity (M)

PWs, IP and MPLS are proven technologies with wide variety of deployments and years of operational experience. Furthermore, the estimated work for missing functionality (packet replication and elimination) does not appear to be extensive, since the existing

protection mechanism already get close to what is needed from the deterministic networking data plane solution.

#### 5.2.3.3. Summary

PseudoWires appear to be a strong candidate as the deterministic networking data plane solution alternative for the DetNet Service layer. The strong points are the technical maturity and the extensive control plane for OAM. This holds specifically for MPLS-based PSN.

Extensions are required to realize the packet replication and duplicate detection features of the deterministic networking data plane.

### 5.2.4. MPLS-Based Ethernet VPN (EVPN)

#### 5.2.4.1. Solution description

MPLS-Based Ethernet VPN (EVPN), in the form documented in [RFC7432] and [RFC7209], is an increasingly popular approach to delivering MPLS-based Ethernet services and is designed to be the successor to Virtual Private LAN Service (VPLS), [RFC4664].

EVPN provides client adaptation and reuses the MPLS data plane discussed above in Section 5.2.2. While not required, the PW Control Word is also used. EVPN control is via BGP, [RFC7432], and may use TE-LSPs, e.g., controlled via [RFC3209] for MPLS transport. Additional EVPN related RFCs and in progress drafts are being developed by the BGP Enabled Services Working Group [6].

#### 5.2.4.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

EVPN generally uses a single MPLS label stack entry to support its client adaptation service. The optional addition of a second label is also supported. In certain cases PW Control Word may also be used.

#2 Flow identification (W)

EVPN currently uses labels to identify flows per {Ethernet Segment Identifier, VLAN} or per MAC level. Additional definition will be needed to standardize identification of finer granularity DetNet flows as well as mapping of TSN services to DetNet Services.

#3 Packet sequencing (M)

Like MPLS, EVPN generally orders packets similar to Ethernet. Reordering is possible primarily during path changes and protection switching. In order to avoid misordering due to ECMP, EVPN uses the "Preferred PW MPLS Control Word" [RFC4385] (in which case EVPN inherits this function from PWs) or the entropy labels [RFC6790].

If additional ordering mechanisms are required, such mechanisms will need to be defined.

#5 Packet replication and elimination (M/W)

EVPN relies on the MPLS layer for all protection functions. See Section 5.1.3 and Section 5.2.2. Some extensions, either at the EVPN or MPLS levels, will be need to support those DetNet applications which require true hitless (i.e., zero loss) 1+1 protection switching. (Network coding may be an interesting alternative to investigate to delivering such hitless loss protection capability.)

#6 Operations, Administration and Maintenance (M/W)

Nodes supporting EVPN may participate in either or both Ethernet level and MPLS level OAM. It is likely that it may make sense to map or adapt the OAM functions at the different levels, but such has yet to be defined. [RFC6371] provides some useful background on this topic.

#8 Class and quality of service capabilities (M/W)

EVPN is largely silent on the topics of CoS and QoS, but the 802.1 TSN Ethernet and existing MPLS TE mechanisms can be directly used. The inter-working of such is new work and within the scope of DetNet. The existing MPLS mechanisms include both CoS (E-LSPs and L-LSPs) and QoS (dedicated TE LSPs).

#10 Technical maturity (M)

EVPN is a second (or third) generation MPLS-based L2VPN service standard. From a data plane standpoint it makes uses of existing MPLS data plane mechanisms. The mechanisms have been widely implemented and deployed.

#### 5.2.4.3. Summary

EVPN is the emerging successor to VPLS. EVPN is standardized, implemented and deployed. It makes use of the mature MPLS data plane. While offering a mature and very comprehensive set of features, certain DetNet required features are not fully/directly supported and additional standardization in these areas are needed. Examples include: mapping CoS and QoS; use of labels per DetNet flow, and hitless 1+1 protection.

### 5.2.5. Higher layer header fields

Fields of headers belonging to higher OSI layers can be used to implement functionality that is not provided e.g., by the IPv6 or IPv4 header fields. However, this approach cannot be always applied, e.g., due to encryption. Furthermore, even if this approach is applicable, it requires deep packet inspection from the routers and switches. There are implementation dependent limits how far into the packet the lookup can be done efficiently in the fast path. When encryption is not used, a safe bet is generally between 128 and 256 octets for the maximum lookup depth. Various higher layer protocols can be applied. Some examples are provided here for the sequence numbering feature (Section 4.3).

#### 5.2.5.1. TCP

The TCP header includes a sequence number parameter, which can be applied to detect and eliminate duplicate packets if DetNet Reliability redundancy is used. As the TCP header is right after the IP header, it does not require very deep packet inspection; the 4-byte sequence number is conveyed by bits 32 through 63 of the TCP header. In addition to sequencing, the TCP header also contain source and destination port information that can be used for assisting the flow identification.

#### 5.2.5.2. RTP

##### 5.2.5.2.1. Solution Description

RTP is often used to deliver time critical traffic in IP networks. RTP is typically carried on top of UDP/IP [RFC3550]. RTP is also augmented by its own control protocol RTCP, which monitors of the data delivery and provides minimal control and identification functionality. RTCP packets do not carry "media payload". Although both RTP and RTCP are typically used with UDP/IP transport they are designed to be independent of the underlying transport and network layers.

The RTP header includes a 2-byte sequence number, which can be used to detect and eliminate duplicate packets if DetNet Reliability redundancy is used. The sequence number is conveyed by bits 16 through 31 of the RTP header. In addition to the sequence number the RTP header has also timestamp field (bits 32 through 63) that can be useful for time synchronization purposes. Furthermore, the RTP header has also one or more synchronization sources (bits starting from 64) that can potentially be useful for flow identification purposes.

5.2.5.2.2. Analysis and Discussion

#1 Encapsulation and overhead (M)

RTP adds minimum 12 octets of header overhead. Typically 8 octets overhead of UDP header has to be also added, at least in a case when RTP is transported over IP. Although RTCP packets do not contribute to the media payload transport they still consume overall network capacity, since all participants to an RTP session including talkers and multicast session listeners are expected to send RTCP reports.

#2 Flow identification (M)

The RTP header contains a synchronization source (SSRC) identifier. The intent is that no two synchronization sources within the same RTP session has the same SSRC identifier.

#3 Packet sequencing (M)

The RTP header contains a 16 bit sequence number.

#5 Packet replication and elimination (M/W)

RTP has precedence of being used for hitless protection switching [ST20227], which essentially is equivalent to DetNet Reliability. Furthermore, recent work in IETF for RTP stream duplication [RFC7198] as a mechanism to protect media flows from packet loss is again equivalent to Detnet Reliability.

#6 Operations, Administration and Maintenance (M)

RTP has its own control protocol RTCP for (minimal) management and stream monitoring purposes. Existing IP OAM tools can directly leveraged when RTP is deployed over IP transport.

#8 Class and quality of service capabilities (M/W)

TBD. [Editor's note: relies on lower layers to provide CoS/QoS]

#10 Technical maturity (M)

RTP has been deployed and used in large commercial systems for over ten years and can be considered a mature technology.

5.2.5.2.3. Summary

RTP appears to be a good candidate as the deterministic networking data plane solution alternative for the DetNet Service layer. The strong points are the technical maturity and the fact it was designed for transporting time-sensitive payload from the beginning. RTP is specifically well suited to be used with (UDP)/IP transport.

Extensions may be required to realize the packet replication and duplicate detection features of the deterministic networking data plane. However, there is already precedence of similar solutions that could potentially be leveraged [ST20227][RFC7198].

## 6. Summary of data plane alternatives

The following table summarizes the criteria (Section 4) used for the evaluation of data plane options.

Applicability per Alternative

| Item # | Meaning |
|---|---|
| #1 | Encapsulation and overhead |
| #2 | Flow identification |
| #3 | Packet sequencing |
| #4 | Explicit routes |
| #5 | Packet replication and elimination |
| #6 | Operations, Administration and Maintenance |
| #8 | Class and quality of service capabilities |
| #9 | Packet traceability |
| #10 | Technical maturity |

Table 5: Evaluation criteria (#7 obsoleted)

There is no single technology that could meet all the criteria on its own. Distinguishing the DetNet Service and the DetNet Transport, as explained in (Section 3), allows a number of combinations, which can meet most of the criteria. There is no room here to evaluate all

possible combinations. Therefore, only some combinations are highlighted here, which are selected based on the number of criteria that are met and the maturity of the technology (#10).

The following table summarizes the evaluation of the data plane options that can be used for the DetNet Transport Layer against the evaluation criteria. Each value in the table is from the corresponding section.

Applicability per Transport Alternative

| Solution | #1 | #2 | #4 | #5 | #6 | #8 | #9 | #10 |
|---|---|---|---|---|---|---|---|---|
| IPv6 | M | W | W | W | M | W | W | M/W |
| IPv4 | M | W | W | W/N | M | M/W | W | M/W |
| MPLS | M | M | M | M/W | M | M/W | M | M |
| BIER | M | M | M | M/W | M/W | M/W | M | W |
| BIER-TE | W/M | N | N | M/W | W | N | W | W |

Summarizing Transport capabilities

Table 6: DetNet Transport Layer

The following table summarizes the evaluation of the data plane options that can be used for the DetNet Service Layer against the criteria evaluation criteria. Each value in the table is from the corresponding section.

Applicability per Service Alternative

| Solution | #1 | #2 | #3 | #5 | #6 | #8 | #10 |
|---|---|---|---|---|---|---|---|
| GRE | M | W | M | W/N | M | W | M |
| PWE3 | M | M | M | W | M/W | M/W | M |
| EVPN | M | W | M | M/W | M/W | M/W | M |
| RTP | M | M | M | M/W | M | M/W | M |

Summarizing Service capabilities

Table 7: DetNet Service Layer

PseudoWire (Section 5.2.3) is the technology that is mature and meets most of the criteria for the DetNet Service layer as shown in the table above. From upper layer protocols PWs or RTP can be a

candidate for non-MPLS PSNs. The identified work for PWs is to figure out how to implement duplicate detection for these protocols (e.g., based on [RFC3985]). In a case of RTP there is precedence of implementing packet duplication and duplicate elimination [ST20227][RFC7198].

PWs can be carried over MPLS or IP. MPLS is the most common technology that is used as PSN for PseudoWires; furthermore, MPLS is a mature technology and meets most DetNet Transport layer criteria. IPv[46] can be also used as PSN and both are mature technologies, although both generally only support CoS (DiffServ) in deployed networks. RTP is independent of the underlying transport technology and network. However, it is well suited for UDP/IP transport.

## 7. Security considerations

This document does not add any new security considerations beyond what the referenced technologies already have.

## 8. IANA Considerations

This document has no IANA considerations.

## 9. Acknowledgements

The author(s) ACK and NACK.

The following people were part of the DetNet Data Plane Design Team:

Jouni Korhonen
Janos Farkas
Norman Finn
Olivier Marce
Gregory Mirsky
Pascal Thubert
Zhuangyan Zhuang

Substantial contributions were received from:

Balazs Varga (service model)

The DetNet chairs serving during the DetNet Data Plane Design Team:

Lou Berger
Pat Thaler

## 10. References

### 10.1. Informative References

[I-D.eckert-bier-te-arch]
Eckert, T., Cauchie, G., Braun, W., and M. Menth, "Traffic Engineering for Bit Index Explicit Replication BIER-TE", draft-eckert-bier-te-arch-03 (work in progress), March 2016.

[I-D.finn-detnet-architecture]
Finn, N., Thubert, P., and M. Teener, "Deterministic Networking Architecture", draft-finn-detnet-architecture-05 (work in progress), June 2016.

[I-D.ietf-6man-rfc2460bis]
Deering, D. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", draft-ietf-6man-rfc2460bis-05 (work in progress), June 2016.

[I-D.ietf-6man-segment-routing-header]
Previdi, S., Filsfils, C., Field, B., Leung, I., Linkova, J., Aries, E., Kosugi, T., Vyncke, E., and D. Lebrun, "IPv6 Segment Routing Header (SRH)", draft-ietf-6man-segment-routing-header-01 (work in progress), March 2016.

[I-D.ietf-bier-architecture]
Wijnands, I., Rosen, E., Dolganow, A., Przygienda, T., and S. Aldrin, "Multicast using Bit Index Explicit Replication", draft-ietf-bier-architecture-03 (work in progress), January 2016.

[I-D.ietf-detnet-problem-statement]
Finn, N. and P. Thubert, "Deterministic Networking Problem Statement", draft-ietf-detnet-problem-statement-00 (work in progress), April 2016.

[I-D.ietf-mpls-residence-time]
Mirsky, G., Ruffini, S., Gray, E., Drake, J., Bryant, S., and S. Vainshtein, "Residence Time Measurement in MPLS network", draft-ietf-mpls-residence-time-10 (work in progress), July 2016.

[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-09 (work in progress), July 2016.

[I-D.ietf-sunset4-gapanalysis]
Perreault, S., Tsou, T., Zhou, C., and P. Fan, "Gap Analysis for IPv4 Sunset", draft-ietf-sunset4-gapanalysis-07 (work in progress), April 2015.

[I-D.kumarzheng-bier-ping]
Kumar, N., Pignataro, C., Akiya, N., Zheng, L., Chen, M., and G. Mirsky, "BIER Ping and Trace", draft-kumarzheng-bier-ping-03 (work in progress), July 2016.

[I-D.mirsky-bier-path-mtu-discovery]
Mirsky, G., Przygienda, T., and A. Dolganow, "Path Maximum Transmission Unit Discovery (PMTUD) for Bit Index Explicit Replication (BIER) Layer", draft-mirsky-bier-path-mtu-discovery-01 (work in progress), April 2016.

[I-D.mirsky-bier-pmmm-oam]
Mirsky, G., Zheng, L., Chen, M., and G. Fioccola, "Performance Measurement (PM) with Marking Method in Bit Index Explicit Replication (BIER) Layer", draft-mirsky-bier-pmmm-oam-01 (work in progress), March 2016.

[I-D.ooamdt-rtgwg-oam-gap-analysis]
Mirsky, G., Nordmark, E., Pignataro, C., Kumar, N., Kumar, D., Chen, M., Mozes, D., and J. Networks, "Operations, Administration and Maintenance (OAM) for Overlay Networks: Gap Analysis", draft-ooamdt-rtgwg-oam-gap-analysis-01 (work in progress), March 2016.

[I-D.ooamdt-rtgwg-ooam-requirement]
Kumar, N., Pignataro, C., Kumar, D., Mirsky, G., Chen, M., Nordmark, E., Networks, J., and D. Mozes, "Overlay OAM Requirements", draft-ooamdt-rtgwg-ooam-requirement-01 (work in progress), July 2016.

[IEEE802.1Qbv]
IEEE, "Enhancements for Scheduled Traffic", 2016, <http://www.ieee802.org/1/files/private/bv-drafts/>.

[IEEE802.1Qca]
IEEE 802.1, "IEEE 802.1Qca Bridges and Bridged Networks - Amendment 24: Path Control and Reservation", IEEE P802.1Qca/D2.1 P802.1Qca, June 2015, <https://standards.ieee.org/findstds/standard/802.1Qca-2015.html>.

[IEEE802.1Qch]
IEEE, "Cyclic Queuing and Forwarding", 2016, <http://www.ieee802.org/1/files/private/ch-drafts/>.

[IEEE8021CB]
Finn, N., "Draft Standard for Local and metropolitan area networks - Seamless Redundancy", IEEE P802.1CB /D2.1 P802.1CB, December 2015, <http://www.ieee802.org/1/files/private/cb-drafts/d2/802-1CB-d2-1.pdf>.

[RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, DOI 10.17487/RFC0791, September 1981, <http://www.rfc-editor.org/info/rfc791>.

[RFC1122] Braden, R., Ed., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, DOI 10.17487/RFC1122, October 1989, <http://www.rfc-editor.org/info/rfc1122>.

[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>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[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, <http://www.rfc-editor.org/info/rfc2474>.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, DOI 10.17487/RFC2784, March 2000, <http://www.rfc-editor.org/info/rfc2784>.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, DOI 10.17487/RFC2890, September 2000, <http://www.rfc-editor.org/info/rfc2890>.

[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>.

[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, <http://www.rfc-editor.org/info/rfc3032>.

[RFC3168] Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <http://www.rfc-editor.org/info/rfc3168>.

[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>.

[RFC3232] Reynolds, J., Ed., "Assigned Numbers: RFC 1700 is Replaced by an On-line Database", RFC 3232, DOI 10.17487/RFC3232, January 2002, <http://www.rfc-editor.org/info/rfc3232>.

[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, <http://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, <http://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, <http://www.rfc-editor.org/info/rfc3473>.

[RFC3550] Schulzrinne, H., Casner, S., Frederick, R., and V. Jacobson, "RTP: A Transport Protocol for Real-Time Applications", STD 64, RFC 3550, DOI 10.17487/RFC3550, July 2003, <http://www.rfc-editor.org/info/rfc3550>.

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <http://www.rfc-editor.org/info/rfc3985>.

[RFC4023] Worster, T., Rekhter, Y., and E. Rosen, Ed., "Encapsulating MPLS in IP or Generic Routing Encapsulation (GRE)", RFC 4023, DOI 10.17487/RFC4023, March 2005, <http://www.rfc-editor.org/info/rfc4023>.

[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, <http://www.rfc-editor.org/info/rfc4385>.

[RFC4446] Martini, L., "IANA Allocations for Pseudowire Edge to Edge Emulation (PWE3)", BCP 116, RFC 4446, DOI 10.17487/RFC4446, April 2006, <http://www.rfc-editor.org/info/rfc4446>.

[RFC4447] Martini, L., Ed., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, DOI 10.17487/RFC4447, April 2006, <http://www.rfc-editor.org/info/rfc4447>.

[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, <http://www.rfc-editor.org/info/rfc4448>.

[RFC4664] Andersson, L., Ed. and E. Rosen, Ed., "Framework for Layer 2 Virtual Private Networks (L2VPNs)", RFC 4664, DOI 10.17487/RFC4664, September 2006, <http://www.rfc-editor.org/info/rfc4664>.

[RFC4817] Townsley, M., Pignataro, C., Wainner, S., Seely, T., and J. Young, "Encapsulation of MPLS over Layer 2 Tunneling Protocol Version 3", RFC 4817, DOI 10.17487/RFC4817, March 2007, <http://www.rfc-editor.org/info/rfc4817>.

[RFC4875] Aggarwal, R., Ed., Papadimitriou, D., Ed., and S. Yasukawa, Ed., "Extensions to Resource Reservation Protocol - Traffic Engineering (RSVP-TE) for Point-to-Multipoint TE Label Switched Paths (LSPs)", RFC 4875, DOI 10.17487/RFC4875, May 2007, <http://www.rfc-editor.org/info/rfc4875>.

[RFC5087] Stein, Y(J)., Shashoua, R., Insler, R., and M. Anavi, "Time Division Multiplexing over IP (TDMoIP)", RFC 5087, DOI 10.17487/RFC5087, December 2007, <http://www.rfc-editor.org/info/rfc5087>.

[RFC5129] Davie, B., Briscoe, B., and J. Tay, "Explicit Congestion Marking in MPLS", RFC 5129, DOI 10.17487/RFC5129, January 2008, <http://www.rfc-editor.org/info/rfc5129>.

[RFC5254] Bitar, N., Ed., Bocci, M., Ed., and L. Martini, Ed., "Requirements for Multi-Segment Pseudowire Emulation Edge-to-Edge (PWE3)", RFC 5254, DOI 10.17487/RFC5254, October 2008, <http://www.rfc-editor.org/info/rfc5254>.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, DOI 10.17487/RFC5305, October 2008, <http://www.rfc-editor.org/info/rfc5305>.

[RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream Label Assignment and Context-Specific Label Space", RFC 5331, DOI 10.17487/RFC5331, August 2008, <http://www.rfc-editor.org/info/rfc5331>.

[RFC5332] Eckert, T., Rosen, E., Ed., Aggarwal, R., and Y. Rekhter, "MPLS Multicast Encapsulations", RFC 5332, DOI 10.17487/RFC5332, August 2008, <http://www.rfc-editor.org/info/rfc5332>.

[RFC5440] Vasseur, JP., Ed. and JL. Le Roux, Ed., "Path Computation Element (PCE) Communication Protocol (PCEP)", RFC 5440, DOI 10.17487/RFC5440, March 2009, <http://www.rfc-editor.org/info/rfc5440>.

[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, <http://www.rfc-editor.org/info/rfc5462>.

[RFC5586] Bocci, M., Ed., Vigoureux, M., Ed., and S. Bryant, Ed., "MPLS Generic Associated Channel", RFC 5586, DOI 10.17487/RFC5586, June 2009, <http://www.rfc-editor.org/info/rfc5586>.

[RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-Segment Pseudowire Emulation Edge-to-Edge", RFC 5659, DOI 10.17487/RFC5659, October 2009, <http://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, <http://www.rfc-editor.org/info/rfc5921>.

[RFC5960] Frost, D., Ed., Bryant, S., Ed., and M. Bocci, Ed., "MPLS Transport Profile Data Plane Architecture", RFC 5960, DOI 10.17487/RFC5960, August 2010, <http://www.rfc-editor.org/info/rfc5960>.

[RFC6275] Perkins, C., Ed., Johnson, D., and J. Arkko, "Mobility Support in IPv6", RFC 6275, DOI 10.17487/RFC6275, July 2011, <http://www.rfc-editor.org/info/rfc6275>.

[RFC6371] Busi, I., Ed. and D. Allan, Ed., "Operations, Administration, and Maintenance Framework for MPLS-Based Transport Networks", RFC 6371, DOI 10.17487/RFC6371, September 2011, <http://www.rfc-editor.org/info/rfc6371>.

[RFC6373] Andersson, L., Ed., Berger, L., Ed., Fang, L., Ed., Bitar, N., Ed., and E. Gray, Ed., "MPLS Transport Profile (MPLS-TP) Control Plane Framework", RFC 6373, DOI 10.17487/RFC6373, September 2011, <http://www.rfc-editor.org/info/rfc6373>.

[RFC6378] Weingarten, Y., Ed., Bryant, S., Osborne, E., Sprecher, N., and A. Fulignoli, Ed., "MPLS Transport Profile (MPLS-TP) Linear Protection", RFC 6378, DOI 10.17487/RFC6378, October 2011, <http://www.rfc-editor.org/info/rfc6378>.

[RFC6426] Gray, E., Bahadur, N., Boutros, S., and R. Aggarwal, "MPLS On-Demand Connectivity Verification and Route Tracing", RFC 6426, DOI 10.17487/RFC6426, November 2011, <http://www.rfc-editor.org/info/rfc6426>.

[RFC6437] Amante, S., Carpenter, B., Jiang, S., and J. Rajahalme, "IPv6 Flow Label Specification", RFC 6437, DOI 10.17487/RFC6437, November 2011, <http://www.rfc-editor.org/info/rfc6437>.

[RFC6564] Krishnan, S., Woodyatt, J., Kline, E., Hoagland, J., and M. Bhatia, "A Uniform Format for IPv6 Extension Headers", RFC 6564, DOI 10.17487/RFC6564, April 2012, <http://www.rfc-editor.org/info/rfc6564>.

[RFC6621] Macker, J., Ed., "Simplified Multicast Forwarding", RFC 6621, DOI 10.17487/RFC6621, May 2012, <http://www.rfc-editor.org/info/rfc6621>.

[RFC6658] Bryant, S., Ed., Martini, L., Swallow, G., and A. Malis, "Packet Pseudowire Encapsulation over an MPLS PSN", RFC 6658, DOI 10.17487/RFC6658, July 2012, <http://www.rfc-editor.org/info/rfc6658>.

[RFC6790] Kompella, K., Drake, J., Amante, S., Henderickx, W., and L. Yong, "The Use of Entropy Labels in MPLS Forwarding", RFC 6790, DOI 10.17487/RFC6790, November 2012, <http://www.rfc-editor.org/info/rfc6790>.

[RFC6814] Pignataro, C. and F. Gont, "Formally Deprecating Some IPv4 Options", RFC 6814, DOI 10.17487/RFC6814, November 2012, <http://www.rfc-editor.org/info/rfc6814>.

[RFC6864] Touch, J., "Updated Specification of the IPv4 ID Field", RFC 6864, DOI 10.17487/RFC6864, February 2013, <http://www.rfc-editor.org/info/rfc6864>.

[RFC7045] Carpenter, B. and S. Jiang, "Transmission and Processing of IPv6 Extension Headers", RFC 7045, DOI 10.17487/RFC7045, December 2013, <http://www.rfc-editor.org/info/rfc7045>.

[RFC7198] Begen, A. and C. Perkins, "Duplicating RTP Streams", RFC 7198, DOI 10.17487/RFC7198, April 2014, <http://www.rfc-editor.org/info/rfc7198>.

[RFC7209] Sajassi, A., Aggarwal, R., Uttaro, J., Bitar, N., Henderickx, W., and A. Isaac, "Requirements for Ethernet VPN (EVPN)", RFC 7209, DOI 10.17487/RFC7209, May 2014, <http://www.rfc-editor.org/info/rfc7209>.

[RFC7271] Ryoo, J., Ed., Gray, E., Ed., van Helvoort, H., D'Alessandro, A., Cheung, T., and E. Osborne, "MPLS Transport Profile (MPLS-TP) Linear Protection to Match the Operational Expectations of Synchronous Digital Hierarchy, Optical Transport Network, and Ethernet Transport Network Operators", RFC 7271, DOI 10.17487/RFC7271, June 2014, <http://www.rfc-editor.org/info/rfc7271>.

[RFC7399] Farrel, A. and D. King, "Unanswered Questions in the Path Computation Element Architecture", RFC 7399, DOI 10.17487/RFC7399, October 2014, <http://www.rfc-editor.org/info/rfc7399>.

[RFC7426] Haleplidis, E., Ed., Pentikousis, K., Ed., Denazis, S., Hadi Salim, J., Meyer, D., and O. Koufopavlou, "Software-Defined Networking (SDN): Layers and Architecture Terminology", RFC 7426, DOI 10.17487/RFC7426, January 2015, <http://www.rfc-editor.org/info/rfc7426>.

[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, <http://www.rfc-editor.org/info/rfc7432>.

[RFC7510] Xu, X., Sheth, N., Yong, L., Callon, R., and D. Black, "Encapsulating MPLS in UDP", RFC 7510, DOI 10.17487/RFC7510, April 2015, <http://www.rfc-editor.org/info/rfc7510>.

[RFC7637] Garg, P., Ed. and Y. Wang, Ed., "NVGRE: Network Virtualization Using Generic Routing Encapsulation", RFC 7637, DOI 10.17487/RFC7637, September 2015, <http://www.rfc-editor.org/info/rfc7637>.

[RFC7676] Pignataro, C., Bonica, R., and S. Krishnan, "IPv6 Support for Generic Routing Encapsulation (GRE)", RFC 7676, DOI 10.17487/RFC7676, October 2015, <http://www.rfc-editor.org/info/rfc7676>.

[RFC7752] Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <http://www.rfc-editor.org/info/rfc7752>.

[RFC7813] Farkas, J., Ed., Bragg, N., Unbehagen, P., Parsons, G., Ashwood-Smith, P., and C. Bowers, "IS-IS Path Control and Reservation", RFC 7813, DOI 10.17487/RFC7813, June 2016, <http://www.rfc-editor.org/info/rfc7813>.

[RFC7872] Gont, F., Linkova, J., Chown, T., and W. Liu, "Observations on the Dropping of Packets with IPv6 Extension Headers in the Real World", RFC 7872, DOI 10.17487/RFC7872, June 2016, <http://www.rfc-editor.org/info/rfc7872>.

[ST20227] SMPTE 2022, "Seamless Protection Switching of SMPTE ST 2022 IP Datagrams", ST 2022-7:2013, 2013, <https://www.smpte.org/digital-library>.

[TSNTG] IEEE Standards Association, "IEEE 802.1 Time-Sensitive Networks Task Group", 2013, <http://www.IEEE802.org/1/pages/avbridges.html>.

10.2. URIs

[1] http://6lab.cisco.com/stats/

[2] https://www.google.com/intl/en/ipv6/statistics.html

[3] https://datatracker.ietf.org/wg/spring/charter/

[4] http://www.iana.org/assignments/g-ach-parameters/g-ach-parameters.xhtml

[5] http://ftp.isi.edu/in-notes/iana/assignments/ethernet-numbers

[6] https://tools.ietf.org/wg/bess/

Appendix A. Examples of combined DetNet Service and Transport layers

Authors' Addresses

Jouni Korhonen (editor)
Broadcom
3151 Zanker Road
San Jose, CA 95134
USA

Email: jouni.nospam@gmail.com

Janos Farkas
Ericsson
Konyves Kalman krt. 11/B
Budapest 1097
Hungary

Email: janos.farkas@ericsson.com

Gregory Mirsky
Ericsson

Email: gregory.mirsky@ericsson.com

Pascal Thubert
Cisco

Email: pthubert@cisco.com

Yan Zhuang
Huawei

Email: zhuangyan.zhuang@huawei.com

Lou Berger
LabN Consulting, L.L.C.

Email: lberger@labn.net

DetNet
Internet-Draft
Intended status: Standards Track
Expires: January 9, 2017

N. Finn
P. Thubert
Cisco
M. Johas Teener
Broadcom

July 8, 2016

# Deterministic Networking Architecture
draft-finn-detnet-architecture-06

## Abstract

Deterministic Networking (DetNet) provides a capability to carry specified unicast or multicast data flows for real-time applications with extremely low data loss rates and bounded latency. Techniques used include: 1) reserving data plane resources for individual (or aggregated) DetNet flows in some or all of the intermediate nodes (e.g. bridges or routers) along the path of the flow; 2) providing fixed paths for DetNet flows that do not rapidly change with the network topology; and 3) sequentializing, replicating, tracing and eliminating duplicate packets at various points to ensure delivery of each packet over at least one path. The capabilities can be managed by configuration, or by manual or automatic network management.

## 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 January 9, 2017.

Table of Contents

## 1. Introduction

Deterministic Networking (DetNet) is a service that can be offered by a network to data flows (DetNet flows) that that are limited, at their source, to a maximum data rate specified by that source. DetNet provides these flows extremely low packet loss rates and assured maximum end-to-end delivery latency. This is accomplished by dedicating network resources such as link bandwidth and buffer space to DetNet flows and/or classes of DetNet flows, and by replicating packets along multiple paths. Unused reserved resources are available to non-DetNet packets.

The Deterministic Networking Problem Statement [I-D.ietf-detnet-problem-statement] introduces Deterministic Networking, and Deterministic Networking Use Cases [I-D.ietf-detnet-use-cases] summarizes the need for it.

A goal of DetNet is a converged network in all respects. That is, the presence of DetNet flows does not preclude non-DetNet flows, and the benefits offered DetNet flows should not, except in extreme cases, prevent existing QoS mechanisms from operating in a normal fashion, subject to the bandwidth required for the DetNet flows. A single source-destination pair can trade both DetNet and non-DetNet flows. End systems and applications need not instantiate special interfaces for DetNet flows. Networks are not restricted to certain topologies; connectivity is not restricted. Any application that generates a data flow that can be usefully characterized as having a maximum bandwidth should be able to take advantage of DetNet, as long as the necessary resources can be reserved. Reservations can be made by the application itself, via network management, by an applications controller, or by other means.

Many applications of interest to Deterministic Networking require the ability to synchronize the clocks in end systems to a sub-microsecond accuracy. Some of the queue control techniques defined in Section 4.3 also require time synchronization among relay and transit nodes. The means used to achieve time synchronization are not addressed in this document. DetNet should accommodate various synchronization techniques and profiles that are defined elsewhere to solve exchange time in different market segments.

The present document is an individual contribution, but it is intended by the authors for adoption by the DetNet working group.

## 2. Terminology

### 2.1. Terms used in this document

The following special terms are used in this document in order to avoid the assumption that a given element in the architecture does or does not have Internet Protocol stack, functions as a router, bridge, firewall, or otherwise plays a particular role at Layer-2 or higher.

destination
: An end system capable of receiving a DetNet flow.

DetNet domain
: The portion of a network that is DetNet aware. It includes end systems and other DetNet nodes.

DetNet flow
: A DetNet flow is a sequence of packets to which the DetNet service is to be applied. It can be limited by the source in its maximum packet size and transmission rate, and can thus be provided congestion-free delivery by the network.

DetNet compound flow and DetNet member flow
: A DetNet compound flow is a DetNet flow that has been separated into multiple duplicate DetNet member flows, which are eventually merged back into a single DetNet compound flow, at the DetNet transport layer. "Compound" and "member" are strictly relative to each other, not absolutes; a DetNet compound flow comprising multiple DetNet member flows can, in turn, be a member of a higher-order compound.

DetNet intermediate node
: A DetNet relay node or transit node.

DetNet relay edge node
: An instance of a DetNet relay node that includes a service layer proxy function for DetNet loss prevention (e.g. packet sequencing and/or elimination) for one or more end systems, analogous to a Label Edge Router (LER).

end system
: Commonly called a "host" or "node" in IETF documents, and an "end station" is IEEE 802 documents. End systems of interest to this document are either sources or destinations of DetNet

flows. And end system may or may not be DetNet transport layer aware or DetNet service layer aware.

link
: A connection between two DetNet nodes. It may be composed of a physical link or a sub-network technology that can provide appropriate traffic delivery for DetNet flows.

DetNet node
: A DetNet aware end system, transit node, or relay node. "DetNet" may be omitted in some text.

Detnet relay node
: A DetNet service layer function that interconnects different DetNet transport layer protocols or networks (instances) to perform packet replication and elimination (Section 3.4. A DetNet relay node typically incorporates DetNet transport layer functions as well, in which case it is collocated with a transit node, such as a bridge, a router, a Label Switch Router (LSR), a firewall, or any other system that participates in the DetNet service layer.

reservation
: A trail of configuration between source to destination(s) through transit nodes and subnets associated with a DetNet flow, required to deliver the benefits of DetNet.

DetNet service layer
: The layer at which loss prevention services such as packet sequencing and the elimination part of replication and elimination (Section 3.4) are performed.

source
: An end system capable of sourcing a DetNet flow.

DetNet transit node
: A node operating at the DetNet transport layer, that utilizes link layer and/or network layer switching across multiple links and/or sub-networks to provide paths for DetNet service layer functions. An MPLS LSR is an example of a DetNet transit node.

DetNet transport layer
: The layer that splits and merges Detnet flows for packet replication and elimination (Section 3.4).

### 2.2. IEEE 802 TSN to DetNet dictionary

This section also serves as a dictionary for translating from the terms used by the IEEE 802 Time-Sensitive Networking (TSN) Task Group to those of the DetNet WG.

Listener
: The IEEE 802 term for a destination of a DetNet flow.

relay system
: The IEEE 802 term for a DetNet intermediate node.

Stream
: The IEEE 802 term for a DetNet flow.

Talker
: The IEEE 802 term for the source of a DetNet flow.

## 3. Providing the DetNet Quality of Service

The DetNet Quality of Service can be expressed in terms of:

- Minimum and maximum end-to-end latency from source to destination; timely delivery and jitter avoidance derive from these constraints

- Probability of loss of a packet, under various assumptions as to the operational states of the nodes and links. A derived property is whether it is acceptable to deliver a duplicate packet, which is an inherent risk in highly reliable and/or broadcast transmissions

It is a distinction of DetNet that it is concerned solely with worst-case values for the end-to-end latency. Average, mean, or typical values are of no interest, because they do not affect the ability of a real-time system to perform its tasks. In general, a trivial priority-based queuing scheme will give better average latency to a data flow than DetNet, but of course, the worst-case latency can be essentially unbounded.

Three techniques are used by DetNet to provide these qualities of service:

- Bandwidth reservation and enforcement (Section 3.1).

- Explicit routes (Section 3.2).

- A DetNet loss protection mechanism.

The DetNet techniques are meant to address both of the DetNet QoS requirements (latency and packet loss). Given that DetNet nodes have a finite amount of buffer space, zero congestion loss necessarily results in a maximum end-to-end latency. It also addresses the largest contribution to packet loss, which is buffer congestion.

After congestion, the most important contributions to packet loss are typically from random media errors and equipment failures. Additional mechanisms, such as encoding schemes and/or data replication techniques are needed. The mechanisms employed are constrained by the requirement to meet the users' latency requirements. Packet replication and elimination (Section 3.4) is one possible mechanism to provide DetNet loss protection.

These three techniques can be applied independently, giving eight possible combinations, including none (no DetNet), although some combinations are of wider utility than others. This separation keeps the protocol stack coherent and maximizes interoperability with existing and developing standards in this (IETF) and other Standards Development Organizations. Some examples of typical expected combinations:

- Explicit routes (a) plus packet replication (b) are exactly the techniques employed by [HSR-PRP]. Explicit routes are achieved by limiting the physical topology of the network, and the sequentialization, replication, and duplicate elimination are facilitated by packet tags added at the front or the end of Ethernet frames.

- Zero congestion loss (a) alone is is offered by IEEE 802.1 Audio Video bridging [IEEE802.1BA-2011]. As long as the network suffers no failures, zero congestion loss can be achieved through the use of a reservation protocol (MSRP), shapers in every bridge, and a bit of network calculus.

- Using all three together gives maximum protection.

There are, of course, simpler methods available (and employed, today) to achieve levels of latency and packet loss that are satisfactory for many applications. Prioritization and over-provisioning is one such technique. However, these methods generally work best in the absence of any significant amount of non-critical traffic in the network (if, indeed, such traffic is supported at all), or work only if the critical traffic constitutes only a small portion of the network's theoretical capacity, or work only if all systems are functioning properly, or in the absence of actions by end systems that disrupt the network's operations.

There are any number of methods in use, defined, or in progress for accomplishing each of the above techniques. It is expected that this DetNet Architecture will assist various vendors, users, and/or "vertical" Standards Development Organizations (dedicated to a single industry) to make selections among the available means of implementing DetNet networks.

### 3.1. Zero Congestion Loss

The primary means by which DetNet achieves its QoS assurances is to completely eliminate congestion at an output port as a cause of packet loss. Given that a DetNet flow cannot be throttled, this can be achieved only by the provision of sufficient buffer storage at each hop through the network to ensure that no packets are dropped due to a lack of buffer storage.

Ensuring adequate buffering requires, in turn, that the source, and every intermediate node along the path to the destination (or nearly every node -- see Section 4.2.2) be careful to regulate its output to not exceed the data rate for any DetNet flow, except for brief periods when making up for interfering traffic. Any packet sent ahead of its time potentially adds to the number of buffers required by the next hop, and may thus exceed the resources allocated for a particular DetNet flow.

The low-level mechanisms described in Section 4.3 provide the necessary regulation of transmissions by an end system or intermediate node to ensure zero congestion loss. The reservation of the bandwidth and buffers for a DetNet flow requires the provisioning described in Section 4.9. A DetNet node may have other resources requiring allocation and/or scheduling, that might otherwise be over-subscribed and trigger the rejection of a reservation.

### 3.2. Explicit routes

In networks controlled by typical peer-to-peer protocols such as IEEE 802.1 ISIS bridged networks or IETF OSPF routed networks, a network topology event in one part of the network can impact, at least briefly, the delivery of data in parts of the network remote from the failure or recovery event. Thus, even redundant paths through a network, if controlled by the typical peer-to-peer protocols, do not eliminate the chances of brief losses of contact.

Many real-time networks rely on physical rings or chains of two-port devices, with a relatively simple ring control protocol. This supports redundant paths with a minimum of wiring. As an additional benefit, ring topologies can often utilize different topology management protocols than those used for a mesh network, with a

consequent reduction in the response time to topology changes. Of course, this comes at some cost in terms of increased hop count, and thus latency, for the typical path.

In order to get the advantages of low hop count and still ensure against even very brief losses of connectivity, DetNet employs explicit routes, where the path taken by a given DetNet flow does not change, at least immediately, and likely not at all, in response to network topology events. When combined with a loss prevention mechanism such as packet replication and elimination (Section 3.4), this results in a high likelihood of continuous connectivity. Explicit routes are commonly used in MPLS TE LSPs.

### 3.3. Jitter Reduction

A core objective of DetNet is to enable the convergence of Non-IP networks onto a common network infrastructure. This requires the accurate emulation of currently deployed mission-specific networks, which typically rely on point-to-point analog (e.g. 4-20mA modulation) and serial-digital cables (or buses) for highly reliable, synchronized and jitter-free communications. While the latency of analog transmissions is basically the speed of light, legacy serial links are usually slow (in the order of Kbps) compared to, say, GigE, and some latency is usually acceptable. What is not acceptable is the introduction of excessive jitter, which may, for instance, affect the stability of control systems.

Applications that are designed to operate on serial links usually do not provide services to recover the jitter, because jitter simply does not exists there. Streams of information are expected to be delivered in-order and the precise time of reception influences the processes. In order to converge such existing applications, there is a desire to emulate all properties of the serial cable, such as clock transportation, perfect flow isolation and fixed latency. While minimal jitter (in the form of specifying minimum, as well as maximum, end-to-end latency) is supported by DetNet, there are practical limitations on packet-based networks in this regard. In general, users are encouraged to use, instead of, "do this when you get the packet," a combination of:

- Sub-microsecond time synchronization among all source and destination end systems, and

- Time-of-execution fields in the application packets.

### 3.4. Packet Replication and Elimination

After congestion loss has been eliminated, the most important causes of packet loss are random media and/or memory faults, and equipment failures. Both causes of packet loss can be greatly reduced by apreading the data in a packet over multiple transmissions. One such method is described in this section, which sends the same packets over multiple paths. Other methods, such as ones that use encoding methods to combine the information in multiple packets, may also be applicable. See Section 6.5.

Packet replication and elimination, also known as seamless redundancy [HSR-PRP], or 1+1 hitless protection, is a function of the DetNet service layer. It involves three capabilities:

- Replicating these packets into multiple DetNet member flows and, typically, sending them along at least two different paths to the destination(s), e.g. over the explicit routes of Section 3.2.

- Providing sequencing information, once, at or near the source, to the packets of a DetNet compound flow. This may be done by adding a sequence number or time stamp as part of DetNet, or may be inherent in the packet, e.g. in a transport protocol, or associated to other physical properties such as the precise time (and radio channel) of reception of the packet.

- Eliminating duplicated packets. This may be done at any step along the path to save network resources further down, in particular if multiple Replication points exist. But the most common case is to perform this operation at the very edge of the DetNet network, preferably in or near the receiver.

This function is a "hitless" version of, e.g., the 1+1 linear protection in [RFC6372]. That is, instead of switching from one flow to the other when a failure of a flow is detected, DetNet combines both flows, and performs a packet-by-packet selection of which to discard, based on sequence number.

In the simplest case, this amounts to replicating each packet in a source that has two interfaces, and conveying them through the network, along separate paths, to the similarly dual-homed destinations, that discard the extras. This ensures that one path (with zero congestion loss) remains, even if some intermediate node fails. The sequence numbers can also be used for loss detection and for re-ordering.

Alternatively, Detnet relay nodes in the network can provide replication and elimination facilities at various points in the network, so that multiple failures can be accommodated.

This is shown in the following figure, where the two relay nodes each replicate (R) the DetNet flow on input, sending the DetNet member flows to both the other relay node and to the end system, and eliminate duplicates (E) on the output interface to the right-hand end system. Any one link in the network can fail, and the Detnet compound flow can still get through. Furthermore, two links can fail, as long as they are in different segments of the network.

Packet replication and elimination

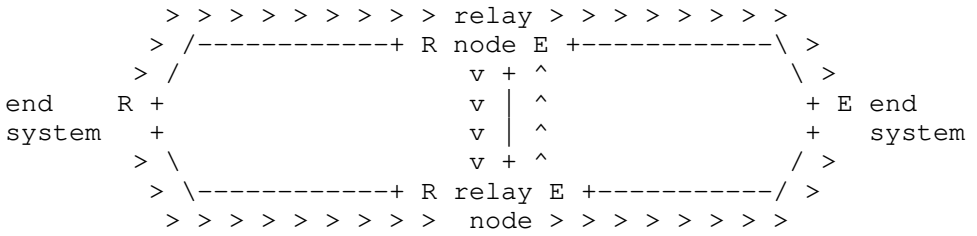



Figure 1

Note that packet replication and elimination does not react to and correct failures; it is entirely passive. Thus, intermittent failures, mistakenly created packet filters, or misrouted data is handled just the same as the equipment failures that are detected handled by typical routing and bridging protocols.

When combining member flows that take different-length paths through the network, and which are also guaranteed a worst-case latency by packet shaping, a merge point may require extra buffering to equalize the delays over the different paths. This equalization ensures that the resultant compound flow will not exceed its contracted bandwidth even after one or the other of the paths is restored after a failure.

# 4. DetNet Architecture

## 4.1. Traffic Engineering for DetNet

Traffic Engineering Architecture and Signaling (TEAS) [TEAS] defines traffic-engineering architectures for generic applicability across packet and non-packet networks. From TEAS perspective, Traffic Engineering (TE) refers to techniques that enable operators to control how specific traffic flows are treated within their networks.

Because if its very nature of establishing explicit optimized paths, Deterministic Networking can be seen as a new, specialized branch of Traffic Engineering, and inherits its architecture with a separation into planes.

The Deterministic Networking architecture is thus composed of three planes, a (User) Application Plane, a Controller Plane, and a Network Plane, which echoes that of Figure 1 of Software-Defined Networking (SDN): Layers and Architecture Terminology [RFC7426].:

#### 4.1.1. The Application Plane

Per [RFC7426], the Application Plane includes both applications and services. In particular, the Application Plane incorporates the User Agent, a specialized application that interacts with the end user / operator and performs requests for Deterministic Networking services via an abstract Flow Management Entity, (FME) which may or may not be collocated with (one of) the end systems.

At the Application Plane, a management interface enables the negotiation of flows between end systems. An abstraction of the flow called a Traffic Specification (TSpec) provides the representation. This abstraction is used to place a reservation over the (Northbound) Service Interface and within the Application plane. It is associated with an abstraction of location, such as IP addresses and DNS names, to identify the end systems and eventually specify intermediate nodes.

#### 4.1.2. The Controller Plane

The Controller Plane corresponds to the aggregation of the Control and Management Planes in [RFC7426], though Common Control and Measurement Plane (CCAMP) [CCAMP] makes an additional distinction between management and measurement. When the logical separation of the Control, Measurement and other Management entities is not relevant, the term Controller Plane is used for simplicity to represent them all, and the term controller refers to any device operating in that plane, whether is it a Path Computation entity or a Network Management entity (NME). The Path Computation Element (PCE) [PCE] is a core element of a controller, in charge of computing Deterministic paths to be applied in the Network Plane.

A (Northbound) Service Interface enables applications in the Application Plane to communicate with the entities in the Controller Plane.

One or more PCE(s) collaborate to implement the requests from the FME as Per-Flow Per-Hop Behaviors installed in the intermediate nodes for

each individual flow. The PCEs place each flow along a deterministic sequence of intermediate nodes so as to respect per-flow constraints such as security and latency, and optimize the overall result for metrics such as an abstract aggregated cost. The deterministic sequence can typically be more complex than a direct sequence and include redundancy path, with one or more packet replication and elimination points.

#### 4.1.3. The Network Plane

The Network Plane represents the network devices and protocols as a whole, regardless of the Layer at which the network devices operate. It includes Forwarding Plane (data plane), Application, and Operational Plane (control plane) aspects.

The network Plane comprises the Network Interface Cards (NIC) in the end systems, which are typically IP hosts, and intermediate nodes, which are typically IP routers and switches. Network-to-Network Interfaces such as used for Traffic Engineering path reservation in [RFC5921], as well as User-to-Network Interfaces (UNI) such as provided by the Local Management Interface (LMI) between network and end systems, are both part of the Network Plane, both in the control plane and the data plane.

A Southbound (Network) Interface enables the entities in the Controller Plane to communicate with devices in the Network Plane. This interface leverages and extends TEAS to describe the physical topology and resources in the Network Plane.

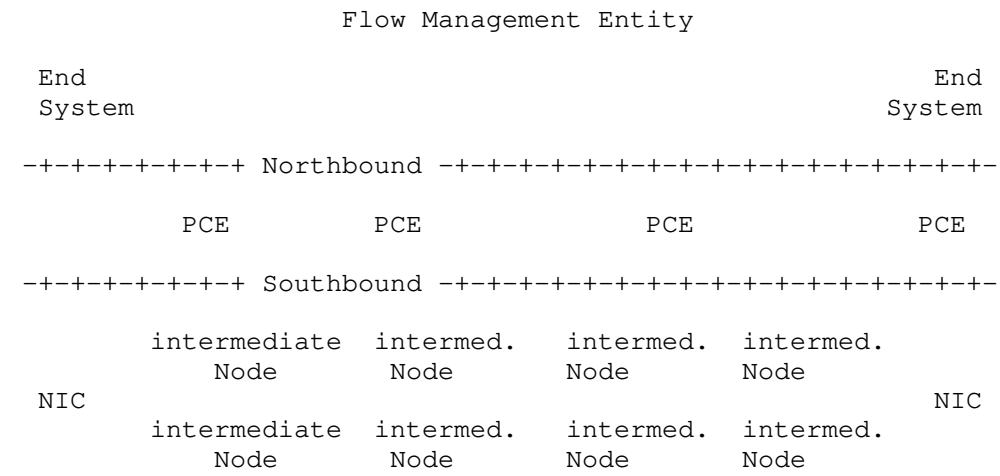



Figure 2

The intermediate nodes (and eventually the end systems NIC) expose their capabilities and physical resources to the controller (the PCE), and update the PCE with their dynamic perception of the topology, across the Southbound Interface. In return, the PCE(s) set the per-flow paths up, providing a Flow Characterization that is more tightly coupled to the intermediate node Operation than a TSpec.

At the Network plane, intermediate nodes may exchange information regarding the state of the paths, between adjacent systems and eventually with the end systems, and forward packets within constraints associated to each flow, or, when unable to do so, perform a last resort operation such as drop or declassify.

This specification focuses on the Southbound interface and the operation of the Network Plane.

## 4.2. DetNet flows

### 4.2.1. Source guarantees

DetNet flows can by synchronous or asynchronous. In synchronous DetNet flows, at least the intermediate nodes (and possibly the end systems) are closely time synchronized, typically to better than 1 microsecond. By transmitting packets from different DetNet flows or classes of DetNet flows at different times, using repeating schedules synchronized among the intermediate nodes, resources such as buffers and link bandwidth can be shared over the time domain among different DetNet flows. There is a tradeoff among techniques for synchronous DetNet flows between the burden of fine-grained scheduling and the benefit of reducing the required resources, especially buffer space.

In contrast, asynchronous DetNet flows are not coordinated with a fine-grained schedule, so relay and end systems must assume worst-case interference among DetNet flows contending for buffer resources. Asynchronous DetNet flows are characterized by:

- A maximum packet size;
- An observation interval; and
- A maximum number of transmissions during that observation interval.

These parameters, together with knowledge of the protocol stack used (and thus the size of the various headers added to a packet), limit the number of bit times per observation interval that the DetNet flow can occupy the physical medium.

The source promises that these limits will not be exceeded. If the source transmits less data than this limit allows, the unused resources such as link bandwidth can be made available by the system to non-DetNet packets. However, making those resources available to DetNet packets in other DetNet flows would serve no purpose. Those other DetNet flows have their own dedicated resources, on the assumption that all DetNet flows can use all of their resources over a long period of time.

Note that there is no provision in DetNet for throttling DetNet flows (reducing the transmission rate via feedback); the assumption is that a DetNet flow, to be useful, must be delivered in its entirety. That is, while any useful application is written to expect a certain number of lost packets, the real-time applications of interest to DetNet demand that the loss of data due to the network is extraordinarily infrequent.

Although DetNet strives to minimize the changes required of an application to allow it to shift from a special-purpose digital network to an Internet Protocol network, one fundamental shift in the behavior of network applications is impossible to avoid: the reservation of resources before the application starts. In the first place, a network cannot deliver finite latency and practically zero packet loss to an arbitrarily high offered load. Secondly, achieving practically zero packet loss for unthrottled (though bandwidth limited) DetNet flows means that bridges and routers have to dedicate buffer resources to specific DetNet flows or to classes of DetNet flows. The requirements of each reservation have to be translated into the parameters that control each system's queuing, shaping, and scheduling functions and delivered to the hosts, bridges, and routers.

#### 4.2.2. Incomplete Networks

The presence in the network of transit nodes or subnets that are not fully capable of offering DetNet services complicates the ability of the intermediate nodes and/or controller to allocate resources, as extra buffering, and thus extra latency, must be allocated at points downstream from the non-DetNet intermediate node for a DetNet flow.

### 4.3. Queuing, Shaping, Scheduling, and Preemption

As described above, DetNet achieves its aims by reserving bandwidth and buffer resources at every hop along the path of the DetNet flow. The reservation itself is not sufficient, however. Implementors and users of a number of proprietary and standard real-time networks have found that standards for specific data plane techniques are required to enable these assurances to be made in a multi-vendor network. The

fundamental reason is that latency variation in one system results in the need for extra buffer space in the next-hop system(s), which in turn, increases the worst-case per-hop latency.

Standard queuing and transmission selection algorithms allow a central controller to compute the latency contribution of each transit node to the end-to-end latency, to compute the amount of buffer space required in each transit node for each incremental DetNet flow, and most importantly, to translate from a flow specification to a set of values for the managed objects that control each relay or end system. The IEEE 802 has specified (and is specifying) a set of queuing, shaping, and scheduling algorithms that enable each transit node (bridge or router), and/or a central controller, to compute these values. These algorithms include:

- A credit-based shaper [IEEE802.1Q-2014] Clause 34.
- Time-gated queues governed by a rotating time schedule, synchronized among all transit nodes [IEEE802.1Qbv].
- Synchronized double (or triple) buffers driven by synchronized time ticks. [IEEE802.1Qch].
- Pre-emption of an Ethernet packet in transmission by a packet with a more stringent latency requirement, followed by the resumption of the preempted packet [IEEE802.1Qbu], [IEEE802.3br].

While these techniques are currently embedded in Ethernet and bridging standards, we can note that they are all, except perhaps for packet preemption, equally applicable to other media than Ethernet, and to routers as well as bridges.

### 4.4. Coexistence with normal traffic

A DetNet network supports the dedication of a high proportion (e.g. 75%) of the network bandwidth to DetNet flows. But, no matter how much is dedicated for DetNet flows, it is a goal of DetNet to coexist with existing Class of Service schemes (e.g., DiffServ). It is also important that non-DetNet traffic not disrupt the DetNet flow, of course (see Section 4.5 and Section 7). For these reasons:

- Bandwidth (transmission opportunities) not utilized by a DetNet flow are available to non-DetNet packets (though not to other DetNet flows).
- DetNet flows can be shaped or scheduled, in order to ensure that the highest-priority non-DetNet packet also is ensured a worst-case latency (at any given hop).

- When transmission opportunities for DetNet flows are scheduled in detail, then the algorithm constructing the schedule should leave sufficient opportunities for non-DetNet packets to satisfy the needs of the users of the network. Detailed scheduling can also permit the time-shared use of buffer resources by different DetNet flows.

Ideally, the net effect of the presence of DetNet flows in a network on the non-DetNet packets is primarily a reduction in the available bandwidth.

### 4.5. Fault Mitigation

One key to building robust real-time systems is to reduce the infinite variety of possible failures to a number that can be analyzed with reasonable confidence. DetNet aids in the process by providing filters and policers to detect DetNet packets received on the wrong interface, or at the wrong time, or in too great a volume, and to then take actions such as discarding the offending packet, shutting down the offending DetNet flow, or shutting down the offending interface.

It is also essential that filters and service remarking be employed at the network edge to prevent non-DetNet packets from being mistaken for DetNet packets, and thus impinging on the resources allocated to DetNet packets.

There exist techniques, at present and/or in various stages of standardization, that can perform these fault mitigation tasks that deliver a high probability that misbehaving systems will have zero impact on well-behaved DetNet flows, except of course, for the receiving interface(s) immediately downstream of the misbehaving device. Examples of such techniques include traffic policing functions (e.g. [RFC2475]) and separating flows into per-flow rate-limited queues.

### 4.6. Representative Protocol Stack Model

Figure 3 illustrates a conceptual DetNet data plane layering model. One may compare it to that in [IEEE802.1CB], Annex C, a work in progress.

DetNet data plane protocol stack

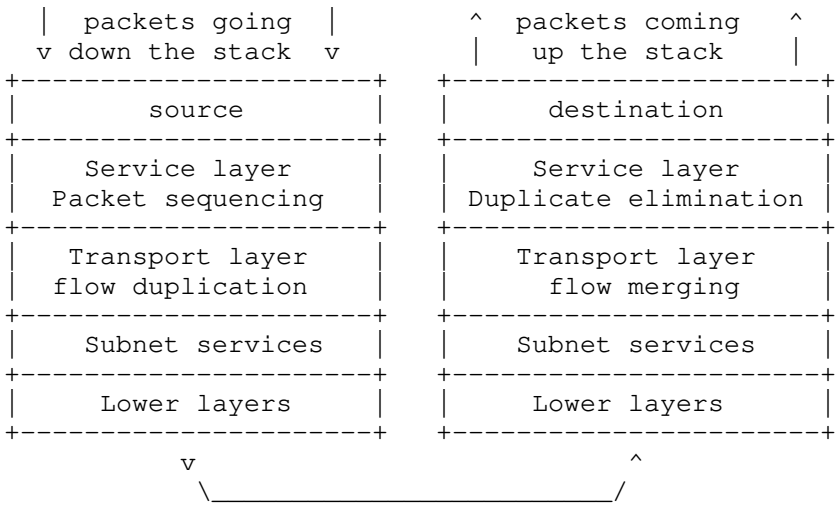



Figure 3

Not all layers are required for any given application, or even for any given network. The layers are, from top to bottom:

Application
: Shown as "source" and "destination" in the diagram.

OAM
: Operations, Administration, and Maintenance leverages in-band and out-of-and signaling that validates whether the service is effectively obtained within QoS constraints. It is not shown in Figure 3; OAM may reside in any number of the layers. OAM can involve specific tagging added in the packets for tracing implementation or network configuration errors; traceability enables to find whether a packet is a replica, which relay node performed the replication, and which segment was intended for the replica.

Packet sequencing
: As part of DetNet loss prevention, supplies the sequence number for providing DetNet loss prevention via packet replication and elimination (Section 3.4) for packets going down the stack. Peers with Duplicate elimination. This layer is not needed if a higher-layer transport protocol is expected to perform any packet sequencing and duplicate elimination required by the DetNet flow duplication.

Duplicate elimination
: As part of the DetNet service layer, based on the sequenced number supplied by its peer, packet sequencing, Duplicate elimination discards any duplicate packets generated by DetNet flow duplication. It can operate on member flows, compound flows, or both. The duplication may also be inferred from other information such as the precise time of reception in a scheduled network. The duplicate elimination layer may also perform resequencing of packets to restore packet order in a flow that was disrupted by the loss of packets on one or another of the multiple paths taken.

Network flow duplication
: As part of DetNet loss prevention, replicates packets going down the stack, that belong to a DetNet compound flow, into two or more DetNet member flows. Note that this function is separate from packet sequencing. Flow duplication can be an explicit duplication and remarking of packets, or can be performed by, for example, techniques similar to ordinary multicast replication. Peers with DetNet flow merging.

Network flow merging
: As part of the DetNet network layer, merges DetNet member flows together for packets coming up the stack belonging to a specific DetNet compound flow. Peers with DetNet flow duplication. DetNet flow merging, together with packet sequencing, duplicate elimination, and DetNet flow duplication, performs packet replication and elimination (Section 3.4).

Queuing shaping scheduling
: The subnet services layer provides the latency and congestion loss parts of the DetNet QoS. See Section 4.3. Note that additional shaping elements may be provided for DetNet edge nodes in order to precondition potentially malformed DetNet flows from a source end system. Note also that these subnet services are typically required of DetNet intermediate nodes that are connected by direct links, not just those connected by subnets such as bridged LANs.

## 4.7. Exporting flow identification

An interesting feature of DetNet, and one that invites implementations that can be accused of "layering violations", is the need for lower layers to be aware of specific flows at higher layers, in order to provide specific queuing and shaping services for specific flows. For example:

- A non-IP, strictly L2 source end system X may be sending multiple flows to the same L2 destination end system Y. Those flows may include DetNet flows with different QoS requirements, and may include non-DetNet flows.

- A router may be sending any number of flows to another router. Again, those flows may include DetNet flows with different QoS requirements, and may include non-DetNet flows.

- Two routers may be separated by bridges. For these bridges to perform any required per-flow queuing and shaping, they must be able to identify the individual flows.

- A Label Edge Router (LERs) may have a Label Switched Path (LSP) set up for handling traffic destined for a particular IP address carrying only non-DetNet flows. If a DetNet flow to that same address is requested, a separate LSP may be needed, in order that all of the Label Switch Routers (LSRs) along the path to the destination give that flow special queuing and shaping.

The need for a lower-level DetNet node to be aware of individual higher-layer flows is not unique to DetNet. But, given the endless complexity of layering and relayering over tunnels that is available to network designers, DetNet needs to provide a model for flow identification that is at least somewhat better than deep packet inspection. That is not to say that deep inspection will not be used, or the capability standardized; but, there are alternatives.

The main alternative is the sequence encode/decode and, particularly, the DetNet flow encoding/decoding layers shown in Figure 3. In this model, at the time a DetNet flow is established and the resources for it reserved, an alternate encapsulation of the DetNet flow at the lower layer is requested and established. For example:

- A single unicast DetNet flow passing from router A through a bridged network to router B may be assigned a {VLAN, multicast destination MAC address} pair that is unique within that bridged network. The bridges can then identify the flow without accessing higher-layer headers. Of course, the receiving router must recognize and accept that multicast MAC address.

- A DetNet flow passing from LSR A to LSR B may be assigned a different label than that used for other flows to the same IP destination.

The DetNet flow encoding/decoding layers shown in Figure 3 perform the required alternate encapsulations. For example, one could place a DetNet flow encoding shim between the Address Resolution Protocol

(ARP) layer and the MAC layer, which alters the {VLAN, MAC address} pair to identify particular streams going up and down the stack, so that the layers above the shim need no alteration to service DetNet flows.

In any of the above cases, it is possible that an existing DetNet flow can be used as a carrier for multiple DetNet sub-flows. (Not to be confused with DetNet compound vs. member flows.) Of course, this requires that the aggregate DetNet flow be provisioned properly to carry the sub-flows.

Thus, rather than deep packet inspection, there is the option to export higher-layer information to the lower layer. The requirement to support one or the other method for flow identification (or both) is the essential complexity that DetNet brings to existing control plane models.

### 4.8. Advertising resources, capabilities and adjacencies

There are three classes of information that a central controller or decentralized control plane needs to know that can only be obtained from the end systems and/or transit nodes in the network. When using a peer-to-peer control plane, some of this information may be required by a system's neighbors in the network.

- Details of the system's capabilities that are required in order to accurately allocate that system's resources, as well as other systems' resources. This includes, for example, which specific queuing and shaping algorithms are implemented (Section 4.3), the number of buffers dedicated for DetNet allocation, and the worst-case forwarding delay.

- The dynamic state of an end or transit node's DetNet resources.

- The identity of the system's neighbors, and the characteristics of the link(s) between the systems, including the length (in nanoseconds) of the link(s).

### 4.9. Provisioning model

#### 4.9.1. Centralized Path Computation and Installation

A centralized routing model, such as provided with a PCE (RFC 4655 [RFC4655]), enables global and per-flow optimizations. (See Section 4.1.) The model is attractive but a number of issues are left to be solved. In particular:

- Whether and how the path computation can be installed by 1) an end device or 2) a Network Management entity,

- And how the path is set up, either by installing state at each hop with a direct interaction between the forwarding device and the PCE, or along a path by injecting a source-routed request at one end of the path.

#### 4.9.2. Distributed Path Setup

Significant work on distributed path setup can be leveraged from MPLS Traffic Engineering, in both its GMPLS and non-GMPLS forms. The protocols within scope are Resource ReSerVation Protocol [RFC3209] [RFC3473](RSVP-TE), OSPF-TE [RFC4203] [RFC5392] and ISIS-TE [RFC5307] [RFC5316]. These should be viewed as starting points as there are feature specific extensions defined that may be applicable to DetNet.

In a Layer-2 only environment, or as part of a layered approach to a mixed environment, IEEE 802.1 also has work, either completed or in progress. [IEEE802.1Q-2014] Clause 35 describes SRP, a peer-to-peer protocol for Layer-2 roughly analogous to RSVP [RFC2205]. [IEEE802.1Qca] defines how ISIS can provide multiple disjoint paths or distribution trees. Also in progress is [IEEE802.1Qcc], which expands the capabilities of SRP.

The integration/interaction of the DetNet control layer with an underlying IEEE 802.1 sub-network control layer will need to be defined.

### 4.10. Scaling to larger networks

Reservations for individual DetNet flows require considerable state information in each transit node, especially when adequate fault mitigation (Section 4.5) is required. The DetNet data plane, in order to support larger numbers of DetNet flows, must support the aggregation of DetNet flows into tunnels, which themselves can be viewed by the transit nodes' data planes largely as individual DetNet flows. Without such aggregation, the per-relay system may limit the scale of DetNet networks.

### 4.11. Connected islands vs. networks

Given that users have deployed examples of the IEEE 802.1 TSN TG standards, which provide capabilities similar to DetNet, it is obvious to ask whether the IETF DetNet effort can be limited to providing Layer-2 connections (VPNs) between islands of bridged TSN networks. While this capability is certainly useful to some applications, and must not be precluded by DetNet, tunneling alone is

not a sufficient goal for the DetNet WG. As shown in the Deterministic Networking Use Cases draft [I-D.ietf-detnet-use-cases], there are already deployments of Layer-2 TSN networks that are encountering the well-known problems of over-large broadcast domains. Routed solutions, and combinations routed/bridged solutions, are both required.

## 5. Compatibility with Layer-2

Standards providing similar capabilities for bridged networks (only) have been and are being generated in the IEEE 802 LAN/MAN Standards Committee. The present architecture describes an abstract model that can be applicable both at Layer-2 and Layer-3, and over links not defined by IEEE 802. It is the intention of the authors (and hopefully, as this draft progresses, of the DetNet Working Group) that IETF and IEEE 802 will coordinate their work, via the participation of common individuals, liaisons, and other means, to maximize the compatibility of their outputs.

DetNet enabled end systems and intermediate nodes can be interconnected by sub-networks, i.e., Layer-2 technologies. These sub-networks will provide DetNet compatible service for support of DetNet traffic. Examples of sub-networks include 802.1TSN and a point-to-point OTN link. Of course, multi-layer DetNet systems may be possible too, where one DetNet appears as a sub-network, and provides service to, a higher layer DetNet system.

## 6. Open Questions

There are a number of architectural questions that will have to be resolved before this document can be submitted for publication. Aside from the obvious fact that this present draft is subject to change, there are specific questions to which the authors wish to direct the readers' attention.

### 6.1. DetNet flow identification and sequencing

The techniques to be used for DetNet flow identification must be settled. The following paragraphs provide a snapshot of the authors' opinions at the time of writing. See [I-D.dt-detnet-dp-alt] for a detailed analysis. See also Section 4.7

IEEE 802.1 TSN streams are identified by giving each stream (DetNet flow) a {VLAN identifier, destination MAC address} pair that is unique in the bridged network, and that the MAC address must be a multicast address. If a source is generating, for example, two unicast UDP flows to the same destination, one DetNet and one not, the DetNet flow's packets must be transformed at some point to have a

multicast destination MAC address, and perhaps, a different VLAN than the non-DetNet flow's packets.

A similar provision would apply to DetNet packets that are identified by MPLS labels; any bridges between the LSRs need a {VLAN identifier, destination MAC address} pair uniquely identifying the DetNet flow in the bridged network.

Provision is made in current draft of [IEEE802.1CB] to make these transformations either in a Layer-2 shim in the source end system, on the output side of a router or LSR, or in a proxy function in the first-hop bridge. It remains to be seen whether this provision is adequate and/or acceptable to the IETF DetNet WG.

There are also questions regarding the sequentialization of packets for use with packet replication and elimination (Section 3.4). [IEEE802.1CB] defines an EtherNet tag carrying a sequence number. If MPLS Pseudowires are used with a control word containing a sequence number, the relationship and interworking between these two formats must be defined.

## 6.2. Flat vs. hierarchical control

Boxes that are solely routers or solely bridges are rare in today's market. In a multi-tenant data center, multiple users' virtual Layer-2/Layer-3 topologies exist simultaneously, implemented on a network whose physical topology bears only accidental resemblance to the virtual topologies.

While the forwarding topology (the bridges and routers) are an important consideration for a DetNet Flow Management Entity (Section 4.1.1), so is the purely physical topology. Ultimately, the model used by the management entities is based on boxes, queues, and links. The authors hope that the work of the TEAS WG will help to clarify exactly what model parameters need to be traded between the intermediate nodes and the controller(s).

## 6.3. Peer-to-peer reservation protocol

As described in Section 4.9.2, the DetNet WG needs to decide whether to support a peer-to-peer protocol for a source and a destination to reserve resources for a DetNet stream. Assuming that enabling the involvement of the source and/or destination is desirable (see Deterministic Networking Use Cases [I-D.ietf-detnet-use-cases]), it remains to decide whether the DetNet WG will make it possible to deploy at least some DetNet capabilities in a network using only a peer-to-peer protocol, without a central controller.

(Note that a UNI (see Section 4.1.3) between an end system and an edge node, for sources and/or listeners to request DetNet services, can be either the first hop of a per-to-peer reservation protocol, or can be deflected by the edge node to a central controller for resolution. Similarly, a decision by a central controller can be effected by the controller instructing the end system or edge node to initiate a per-to-peer protocol activity.)

### 6.4. Wireless media interactions

Deterministic Networking Use Cases [I-D.ietf-detnet-use-cases] illustrates cases where wireless media are needed in a DetNet network. Some wireless media in general use, such as IEEE 802.11 [IEEE802.1Q-2014], have significantly higher packet loss rates than typical wired media, such as Ethernet [IEEE802.3-2012]. IEEE 802.11 includes support for such features as MAC-layer acknowledgements and retransmissions.

The techniques described in Section 3 are likely to improve the ability of a mixed wired/wireless network to offer the DetNet QoS features. The interaction of these techniques with the features of specific wireless media, although they may be significant, cannot be addressed in this document. It remains to be decided to what extent the DetNet WG will address them, and to what extent other WGs, e.g. 6TiSCH, will do so.

### 6.5. Packet encoding for loss prevention

There are other methods for reducing packet loss caused by random hardware errors and/or equipment failues that involve encoding the information in a packet belonging to a DetNet flow into multiple transmission units, typically combining information from multiple packets into any given transmission unit. Such techniques may be applicable for use as a DetNet loss prevention technique, assuming that the DetNet users' needs for timeliness of delivery and freedom from interference with misbehaving DetNet flows can be met.

## 7. Security Considerations

Security in the context of Deterministic Networking has an added dimension; the time of delivery of a packet can be just as important as the contents of the packet, itself. A man-in-the-middle attack, for example, can impose, and then systematically adjust, additional delays into a link, and thus disrupt or subvert a real-time application without having to crack any encryption methods employed. See [RFC7384] for an exploration of this issue in a related context.

Furthermore, in a control system where millions of dollars of equipment, or even human lives, can be lost if the DetNet QoS is not delivered, one must consider not only simple equipment failures, where the box or wire instantly becomes perfectly silent, but bizarre errors such as can be caused by software failures. Because there is essential no limit to the kinds of failures that can occur, protecting against realistic equipment failures is indistinguishable, in most cases, from protecting against malicious behavior, whether accidental or intentional. See also Section 4.5.

Security must cover:

- the protection of the signaling protocol
- the authentication and authorization of the controlling systems
- the identification and shaping of the DetNet flows

## 8. Privacy Considerations

DetNet is provides a Quality of Service (QoS), and as such, does not directly raise any new privacy considerations.

However, the requirement for every (or almost every) node along the path of a DetNet flow to identify DetNet flows may present an additional attack surface for privacy, should the DetNet paradigm be found useful in broader environments.

## 9. IANA Considerations

This document does not require an action from IANA.

## 10. Acknowledgements

The authors wish to thank Jouni Korhonen, Erik Nordmark, George Swallow, Rudy Klecka, Anca Zamfir, David Black, Thomas Watteyne, Shitanshu Shah, Craig Gunther, Rodney Cummings, Balazs Varga, Wilfried Steiner, Marcel Kiessling, Karl Weber, Janos Farkas, Ethan Grossman, Pat Thaler, and Lou Berger for their various contribution with this work.

## 11. Access to IEEE 802.1 documents

To access password protected IEEE 802.1 drafts, see the IETF IEEE 802.1 information page at https://www.ietf.org/proceedings/52/slides/bridge-0/tsld003.htm.

## 12. Informative References

[AVnu] http://www.avnu.org/, "The AVnu Alliance tests and certifies devices for interoperability, providing a simple and reliable networking solution for AV network implementation based on the Audio Video Bridging (AVB) standards.".

[CCAMP] IETF, "Common Control and Measurement Plane", <https://datatracker.ietf.org/doc/charter-ietf-ccamp/>.

[HART] www.hartcomm.org, "Highway Addressable Remote Transducer, a group of specifications for industrial process and control devices administered by the HART Foundation".

[HSR-PRP] IEC, "High availability seamless redundancy (HSR) is a further development of the PRP approach, although HSR functions primarily as a protocol for creating media redundancy while PRP, as described in the previous section, creates network redundancy. PRP and HSR are both described in the IEC 62439 3 standard.", <http://webstore.iec.ch/webstore/webstore.nsf/artnum/046615!opendocument>.

[I-D.dt-detnet-dp-alt] Korhonen, J., Farkas, J., Mirsky, G., Thubert, P., Zhuangyan, Z., and L. Berger, "DetNet Data Plane Protocol and Solution Alternatives", draft-dt-detnet-dp-alt-01 (work in progress), July 2016.

[I-D.ietf-6tisch-architecture] Thubert, P., "An Architecture for IPv6 over the TSCH mode of IEEE 802.15.4", draft-ietf-6tisch-architecture-10 (work in progress), June 2016.

[I-D.ietf-6tisch-tsch] Watteyne, T., Palattella, M., and L. Grieco, "Using IEEE802.15.4e TSCH in an IoT context: Overview, Problem Statement and Goals", draft-ietf-6tisch-tsch-06 (work in progress), March 2015.

[I-D.ietf-detnet-problem-statement] Finn, N. and P. Thubert, "Deterministic Networking Problem Statement", draft-ietf-detnet-problem-statement-00 (work in progress), April 2016.

[I-D.ietf-detnet-use-cases]
Grossman, E., Gunther, C., Thubert, P., Wetterwald, P., Raymond, J., Korhonen, J., Kaneko, Y., Das, S., Zha, Y., Varga, B., Farkas, J., Goetz, F., and J. Schmitt, "Deterministic Networking Use Cases", draft-ietf-detnet-use-cases-10 (work in progress), July 2016.

[I-D.ietf-roll-rpl-industrial-applicability]
Phinney, T., Thubert, P., and R. Assimiti, "RPL applicability in industrial networks", draft-ietf-roll-rpl-industrial-applicability-02 (work in progress), October 2013.

[I-D.svshah-tsvwg-deterministic-forwarding]
Shah, S. and P. Thubert, "Deterministic Forwarding PHB", draft-svshah-tsvwg-deterministic-forwarding-04 (work in progress), August 2015.

[IEEE802.11-2012]
IEEE, "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications", 2012, <http://standards.ieee.org/getieee802/download/802.11-2012.pdf>.

[IEEE802.1AS-2011]
IEEE, "Timing and Synchronizations (IEEE 802.1AS-2011)", 2011, <http://standards.ieee.org/getIEEE802/download/802.1AS-2011.pdf>.

[IEEE802.1BA-2011]
IEEE, "AVB Systems (IEEE 802.1BA-2011)", 2011, <http://standards.ieee.org/getIEEE802/download/802.1BA-2011.pdf>.

[IEEE802.1CB]
IEEE, "Frame Replication and Elimination for Reliability (IEEE Draft P802.1CB)", 2016, <http://www.ieee802.org/1/files/private/cb-drafts/>.

[IEEE802.1Q-2014]
IEEE, "MAC Bridges and VLANs (IEEE 802.1Q-2014", 2014, <http://standards.ieee.org/getieee802/download/802-1Q-2014.pdf>.

[IEEE802.1Qbu]
IEEE, "Frame Preemption", 2016, <http://www.ieee802.org/1/files/private/bu-drafts/>.

[IEEE802.1Qbv]
IEEE, "Enhancements for Scheduled Traffic", 2016, <http://www.ieee802.org/1/files/private/bv-drafts/>.

[IEEE802.1Qca]
IEEE 802.1, "IEEE 802.1Qca Bridges and Bridged Networks - Amendment 24: Path Control and Reservation", IEEE P802.1Qca/D2.1 P802.1Qca, June 2015, <https://standards.ieee.org/findstds/standard/802.1Qca-2015.html>.

[IEEE802.1Qcc]
IEEE, "Stream Reservation Protocol (SRP) Enhancements and Performance Improvements", 2016, <http://www.ieee802.org/1/files/private/cc-drafts/>.

[IEEE802.1Qch]
IEEE, "Cyclic Queuing and Forwarding", 2016, <http://www.ieee802.org/1/files/private/ch-drafts/>.

[IEEE802.1TSNTG]
IEEE Standards Association, "IEEE 802.1 Time-Sensitive Networks Task Group", 2013, <http://www.IEEE802.org/1/pages/avbridges.html>.

[IEEE802.3-2012]
IEEE, "IEEE Stabdard for Ethernet", 2012, <http://standards.ieee.org/getieee802/download/802.3-2012.pdf>.

[IEEE802.3br]
IEEE, "Interspersed Express Traffic", 2016, <http://www.ieee802.org/3/br/>.

[IEEE802154]
IEEE standard for Information Technology, "IEEE std. 802.15.4, Part. 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks", June 2011.

[IEEE802154e]
IEEE standard for Information Technology, "IEEE std. 802.15.4e, Part. 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC sublayer", April 2012.

[ISA100.11a]
ISA/IEC, "ISA100.11a, Wireless Systems for Automation, also IEC 62734", 2011, < http://www.isa100wci.org/en-US/Documents/PDF/3405-ISA100-WirelessSystems-Future-broch-WEB-ETSI.aspx>.

[ISA95] ANSI/ISA, "Enterprise-Control System Integration Part 1: Models and Terminology", 2000, <https://www.isa.org/isa95/>.

[ODVA] http://www.odva.org/, "The organization that supports network technologies built on the Common Industrial Protocol (CIP) including EtherNet/IP.".

[PCE] IETF, "Path Computation Element", <https://datatracker.ietf.org/doc/charter-ietf-pce/>.

[Profinet]
http://us.profinet.com/technology/profinet/, "PROFINET is a standard for industrial networking in automation.", <http://us.profinet.com/technology/profinet/>.

[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>.

[RFC2475] Blake, S., Black, D., Carlson, M., Davies, E., Wang, Z., and W. Weiss, "An Architecture for Differentiated Services", RFC 2475, DOI 10.17487/RFC2475, December 1998, <http://www.rfc-editor.org/info/rfc2475>.

[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>.

[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, <http://www.rfc-editor.org/info/rfc3473>.

[RFC4203] Kompella, K., Ed. and Y. Rekhter, Ed., "OSPF Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 4203, DOI 10.17487/RFC4203, October 2005, <http://www.rfc-editor.org/info/rfc4203>.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation Element (PCE)-Based Architecture", RFC 4655, DOI 10.17487/RFC4655, August 2006, <http://www.rfc-editor.org/info/rfc4655>.

[RFC5307] Kompella, K., Ed. and Y. Rekhter, Ed., "IS-IS Extensions in Support of Generalized Multi-Protocol Label Switching (GMPLS)", RFC 5307, DOI 10.17487/RFC5307, October 2008, <http://www.rfc-editor.org/info/rfc5307>.

[RFC5316] Chen, M., Zhang, R., and X. Duan, "ISIS Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5316, DOI 10.17487/RFC5316, December 2008, <http://www.rfc-editor.org/info/rfc5316>.

[RFC5392] Chen, M., Zhang, R., and X. Duan, "OSPF Extensions in Support of Inter-Autonomous System (AS) MPLS and GMPLS Traffic Engineering", RFC 5392, DOI 10.17487/RFC5392, January 2009, <http://www.rfc-editor.org/info/rfc5392>.

[RFC5673] Pister, K., Ed., Thubert, P., Ed., Dwars, S., and T. Phinney, "Industrial Routing Requirements in Low-Power and Lossy Networks", RFC 5673, DOI 10.17487/RFC5673, October 2009, <http://www.rfc-editor.org/info/rfc5673>.

[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, <http://www.rfc-editor.org/info/rfc5921>.

[RFC6372] Sprecher, N., Ed. and A. Farrel, Ed., "MPLS Transport Profile (MPLS-TP) Survivability Framework", RFC 6372, DOI 10.17487/RFC6372, September 2011, <http://www.rfc-editor.org/info/rfc6372>.

[RFC7384] Mizrahi, T., "Security Requirements of Time Protocols in Packet Switched Networks", RFC 7384, DOI 10.17487/RFC7384, October 2014, <http://www.rfc-editor.org/info/rfc7384>.

[RFC7426] Haleplidis, E., Ed., Pentikousis, K., Ed., Denazis, S., Hadi Salim, J., Meyer, D., and O. Koufopavlou, "Software-Defined Networking (SDN): Layers and Architecture Terminology", RFC 7426, DOI 10.17487/RFC7426, January 2015, <http://www.rfc-editor.org/info/rfc7426>.

[TEAS] IETF, "Traffic Engineering Architecture and Signaling", <https://datatracker.ietf.org/doc/charter-ietf-teas/>.

[WirelessHART]
www.hartcomm.org, "Industrial Communication Networks - Wireless Communication Network and Communication Profiles - WirelessHART - IEC 62591", 2010.

## Authors' Addresses

Norman Finn
Cisco Systems
170 W Tasman Dr.
San Jose, California  95134
USA

Phone: +1 408 526 4495
Email: nfinn@cisco.com

Pascal Thubert
Cisco Systems
Village d'Entreprises Green Side
400, Avenue de Roumanille
Batiment T3
Biot - Sophia Antipolis  06410
FRANCE

Phone: +33 4 97 23 26 34
Email: pthubert@cisco.com

Michael Johas Teener
Broadcom Corp.
3151 Zanker Rd.
San Jose, California  95134
USA

Phone: +1 831 824 4228
Email: MikeJT@broadcom.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: January 5, 2017

E. Grossman, Ed.
DOLBY
C. Gunther
HARMAN
P. Thubert
P. Wetterwald
CISCO
J. Raymond
HYDRO-QUEBEC
J. Korhonen
BROADCOM
Y. Kaneko
Toshiba
S. Das
Applied Communication Sciences
Y. Zha
HUAWEI
B. Varga
J. Farkas
Ericsson
F. Goetz
J. Schmitt
Siemens

July 4, 2016

# Deterministic Networking Use Cases
draft-ietf-detnet-use-cases-10

Abstract

This draft documents requirements in several diverse industries to establish multi-hop paths for characterized flows with deterministic properties. In this context deterministic implies that streams can be established which provide guaranteed bandwidth and latency which can be established from either a Layer 2 or Layer 3 (IP) interface, and which can co-exist on an IP network with best-effort traffic.

Additional requirements include optional redundant paths, very high reliability paths, time synchronization, and clock distribution. Industries considered include wireless for industrial applications, professional audio, electrical utilities, building automation systems, radio/mobile access networks, automotive, and gaming.

For each case, this document will identify the application, identify representative solutions used today, and what new uses an IETF DetNet solution may enable.

Status of This Memo



Copyright Notice



Table of Contents

# 1. Introduction

This draft presents use cases from diverse industries which have in common a need for deterministic streams, but which also differ notably in their network topologies and specific desired behavior. Together, they provide broad industry context for DetNet and a yardstick against which proposed DetNet designs can be measured (to what extent does a proposed design satisfy these various use cases?)

For DetNet, use cases explicitly do not define requirements; The DetNet WG will consider the use cases, decide which elements are in scope for DetNet, and the results will be incorporated into future drafts. Similarly, the DetNet use case draft explicitly does not suggest any specific design, architecture or protocols, which will be topics of future drafts.

We present for each use case the answers to the following questions:

- What is the use case?
- How is it addressed today?
- How would you like it to be addressed in the future?
- What do you want the IETF to deliver?

The level of detail in each use case should be sufficient to express the relevant elements of the use case, but not more.

At the end we consider the use cases collectively, and examine the most significant goals they have in common.

# 2. Pro Audio and Video

## 2.1. Use Case Description

The professional audio and video industry ("ProAV") includes:

- Music and film content creation
- Broadcast
- Cinema
- Live sound

- Public address, media and emergency systems at large venues (airports, stadiums, churches, theme parks).

These industries have already transitioned audio and video signals from analog to digital. However, the digital interconnect systems remain primarily point-to-point with a single (or small number of) signals per link, interconnected with purpose-built hardware.

These industries are now transitioning to packet-based infrastructure to reduce cost, increase routing flexibility, and integrate with existing IT infrastructure.

Today ProAV applications have no way to establish deterministic streams from a standards-based Layer 3 (IP) interface, which is a fundamental limitation to the use cases described here. Today deterministic streams can be created within standards-based layer 2 LANs (e.g. using IEEE 802.1 AVB) however these are not routable via IP and thus are not effective for distribution over wider areas (for example broadcast events that span wide geographical areas).

It would be highly desirable if such streams could be routed over the open Internet, however solutions with more limited scope (e.g. enterprise networks) would still provide a substantial improvement.

The following sections describe specific ProAV use cases.

#### 2.1.1. Uninterrupted Stream Playback

Transmitting audio and video streams for live playback is unlike common file transfer because uninterrupted stream playback in the presence of network errors cannot be achieved by re-trying the transmission; by the time the missing or corrupt packet has been identified it is too late to execute a re-try operation. Buffering can be used to provide enough delay to allow time for one or more retries, however this is not an effective solution in applications where large delays (latencies) are not acceptable (as discussed below).

Streams with guaranteed bandwidth can eliminate congestion on the network as a cause of transmission errors that would lead to playback interruption. Use of redundant paths can further mitigate transmission errors to provide greater stream reliability.

#### 2.1.2. Synchronized Stream Playback

Latency in this context is the time between when a signal is initially sent over a stream and when it is received. A common example in ProAV is time-synchronizing audio and video when they take

separate paths through the playback system. In this case the latency of both the audio and video streams must be bounded and consistent if the sound is to remain matched to the movement in the video. A common tolerance for audio/video sync is one NTSC video frame (about 33ms) and to maintain the audience perception of correct lip sync the latency needs to be consistent within some reasonable tolerance, for example 10%.

A common architecture for synchronizing multiple streams that have different paths through the network (and thus potentially different latencies) is to enable measurement of the latency of each path, and have the data sinks (for example speakers) delay (buffer) all packets on all but the slowest path. Each packet of each stream is assigned a presentation time which is based on the longest required delay. This implies that all sinks must maintain a common time reference of sufficient accuracy, which can be achieved by any of various techniques.

This type of architecture is commonly implemented using a central controller that determines path delays and arbitrates buffering delays.

#### 2.1.3. Sound Reinforcement

Consider the latency (delay) from when a person speaks into a microphone to when their voice emerges from the speaker. If this delay is longer than about 10-15 milliseconds it is noticeable and can make a sound reinforcement system unusable (see slide 6 of [SRP_LATENCY]). (If you have ever tried to speak in the presence of a delayed echo of your voice you may know this experience).

Note that the 15ms latency bound includes all parts of the signal path, not just the network, so the network latency must be significantly less than 15ms.

In some cases local performers must perform in synchrony with a remote broadcast. In such cases the latencies of the broadcast stream and the local performer must be adjusted to match each other, with a worst case of one video frame (33ms for NTSC video).

In cases where audio phase is a consideration, for example beam-forming using multiple speakers, latency requirements can be in the 10 microsecond range (1 audio sample at 96kHz).

#### 2.1.4. Deterministic Time to Establish Streaming

Note: It is still under WG discussion whether this topic (stream startup time) is within scope of DetNet.

Some audio systems installed in public environments (airports, hospitals) have unique requirements with regards to health, safety and fire concerns. One such requirement is a maximum of 3 seconds for a system to respond to an emergency detection and begin sending appropriate warning signals and alarms without human intervention. For this requirement to be met, the system must support a bounded and acceptable time from a notification signal to specific stream establishment. For further details see [ISO7240-16].

Similar requirements apply when the system is restarted after a power cycle, cable re-connection, or system reconfiguration.

In many cases such re-establishment of streaming state must be achieved by the peer devices themselves, i.e. without a central controller (since such a controller may only be present during initial network configuration).

Video systems introduce related requirements, for example when transitioning from one camera feed (video stream) to another (see [STUDIO_IP] and [ESPN_DC2]).

#### 2.1.5. Secure Transmission

##### 2.1.5.1. Safety

Professional audio systems can include amplifiers that are capable of generating hundreds or thousands of watts of audio power which if used incorrectly can cause hearing damage to those in the vicinity. Apart from the usual care required by the systems operators to prevent such incidents, the network traffic that controls these devices must be secured (as with any sensitive application traffic).

### 2.2. Pro Audio Today

Some proprietary systems have been created which enable deterministic streams at Layer 3 however they are "engineered networks" which require careful configuration to operate, often require that the system be over-provisioned, and it is implied that all devices on the network voluntarily play by the rules of that network. To enable these industries to successfully transition to an interoperable multi-vendor packet-based infrastructure requires effective open standards, and we believe that establishing relevant IETF standards is a crucial factor.

## 2.3. Pro Audio Future

### 2.3.1. Layer 3 Interconnecting Layer 2 Islands

It would be valuable to enable IP to connect multiple Layer 2 LANs.

As an example, ESPN recently constructed a state-of-the-art 194,000 sq ft, $125 million broadcast studio called DC2. The DC2 network is capable of handling 46 Tbps of throughput with 60,000 simultaneous signals. Inside the facility are 1,100 miles of fiber feeding four audio control rooms (see [ESPN_DC2] ).

In designing DC2 they replaced as much point-to-point technology as they could with packet-based technology. They constructed seven individual studios using layer 2 LANS (using IEEE 802.1 AVB) that were entirely effective at routing audio within the LANs. However to interconnect these layer 2 LAN islands together they ended up using dedicated paths in a custom SDN (Software Defined Networking) router because there is no standards-based routing solution available.

### 2.3.2. High Reliability Stream Paths

On-air and other live media streams are often backed up with redundant links that seamlessly act to deliver the content when the primary link fails for any reason. In point-to-point systems this is provided by an additional point-to-point link; the analogous requirement in a packet-based system is to provide an alternate path through the network such that no individual link can bring down the system.

### 2.3.3. Integration of Reserved Streams into IT Networks

A commonly cited goal of moving to a packet based media infrastructure is that costs can be reduced by using off the shelf, commodity network hardware. In addition, economy of scale can be realized by combining media infrastructure with IT infrastructure. In keeping with these goals, stream reservation technology should be compatible with existing protocols, and not compromise use of the network for best effort (non-time-sensitive) traffic.

### 2.3.4. Use of Unused Reservations by Best-Effort Traffic

In cases where stream bandwidth is reserved but not currently used (or is under-utilized) that bandwidth must be available to best-effort (i.e. non-time-sensitive) traffic. For example a single stream may be nailed up (reserved) for specific media content that needs to be presented at different times of the day, ensuring timely delivery of that content, yet in between those times the full

bandwidth of the network can be utilized for best-effort tasks such as file transfers.

This also addresses a concern of IT network administrators that are considering adding reserved bandwidth traffic to their networks that ("users will reserve large quantities of bandwidth and then never un-reserve it even though they are not using it, and soon the network will have no bandwidth left").

#### 2.3.5. Traffic Segregation

Note: It is still under WG discussion whether this topic will be addressed by DetNet.

Sink devices may be low cost devices with limited processing power. In order to not overwhelm the CPUs in these devices it is important to limit the amount of traffic that these devices must process.

As an example, consider the use of individual seat speakers in a cinema. These speakers are typically required to be cost reduced since the quantities in a single theater can reach hundreds of seats. Discovery protocols alone in a one thousand seat theater can generate enough broadcast traffic to overwhelm a low powered CPU. Thus an installation like this will benefit greatly from some type of traffic segregation that can define groups of seats to reduce traffic within each group. All seats in the theater must still be able to communicate with a central controller.

There are many techniques that can be used to support this requirement including (but not limited to) the following examples.

##### 2.3.5.1. Packet Forwarding Rules, VLANs and Subnets

Packet forwarding rules can be used to eliminate some extraneous streaming traffic from reaching potentially low powered sink devices, however there may be other types of broadcast traffic that should be eliminated using other means for example VLANs or IP subnets.

##### 2.3.5.2. Multicast Addressing (IPv4 and IPv6)

Multicast addressing is commonly used to keep bandwidth utilization of shared links to a minimum.

Because of the MAC Address forwarding nature of Layer 2 bridges it is important that a multicast MAC address is only associated with one stream. This will prevent reservations from forwarding packets from one stream down a path that has no interested sinks simply because

there is another stream on that same path that shares the same multicast MAC address.

Since each multicast MAC Address can represent 32 different IPv4 multicast addresses there must be a process put in place to make sure this does not occur. Requiring use of IPv6 address can achieve this, however due to their continued prevalence, solutions that are effective for IPv4 installations are also required.

#### 2.3.6. Latency Optimization by a Central Controller

A central network controller might also perform optimizations based on the individual path delays, for example sinks that are closer to the source can inform the controller that they can accept greater latency since they will be buffering packets to match presentation times of farther away sinks. The controller might then move a stream reservation on a short path to a longer path in order to free up bandwidth for other critical streams on that short path. See slides 3-5 of [SRP_LATENCY].

Additional optimization can be achieved in cases where sinks have differing latency requirements, for example in a live outdoor concert the speaker sinks have stricter latency requirements than the recording hardware sinks. See slide 7 of [SRP_LATENCY].

#### 2.3.7. Reduced Device Cost Due To Reduced Buffer Memory

Device cost can be reduced in a system with guaranteed reservations with a small bounded latency due to the reduced requirements for buffering (i.e. memory) on sink devices. For example, a theme park might broadcast a live event across the globe via a layer 3 protocol; in such cases the size of the buffers required is proportional to the latency bounds and jitter caused by delivery, which depends on the worst case segment of the end-to-end network path. For example on todays open internet the latency is typically unacceptable for audio and video streaming without many seconds of buffering. In such scenarios a single gateway device at the local network that receives the feed from the remote site would provide the expensive buffering required to mask the latency and jitter issues associated with long distance delivery. Sink devices in the local location would have no additional buffering requirements, and thus no additional costs, beyond those required for delivery of local content. The sink device would be receiving the identical packets as those sent by the source and would be unaware that there were any latency or jitter issues along the path.

### 2.4. Pro Audio Asks

- Layer 3 routing on top of AVB (and/or other high QoS networks)
- Content delivery with bounded, lowest possible latency
- IntServ and DiffServ integration with AVB (where practical)
- Single network for A/V and IT traffic
- Standards-based, interoperable, multi-vendor
- IT department friendly
- Enterprise-wide networks (e.g. size of San Francisco but not the whole Internet (yet...))

## 3. Electrical Utilities

### 3.1. Use Case Description

Many systems that an electrical utility deploys today rely on high availability and deterministic behavior of the underlying networks. Here we present use cases in Transmission, Generation and Distribution, including key timing and reliability metrics. We also discuss security issues and industry trends which affect the architecture of next generation utility networks

#### 3.1.1. Transmission Use Cases

##### 3.1.1.1. Protection

Protection means not only the protection of human operators but also the protection of the electrical equipment and the preservation of the stability and frequency of the grid. If a fault occurs in the transmission or distribution of electricity then severe damage can occur to human operators, electrical equipment and the grid itself, leading to blackouts.

Communication links in conjunction with protection relays are used to selectively isolate faults on high voltage lines, transformers, reactors and other important electrical equipment. The role of the teleprotection system is to selectively disconnect a faulty part by transferring command signals within the shortest possible time.

#### 3.1.1.1.1. Key Criteria

The key criteria for measuring teleprotection performance are command transmission time, dependability and security. These criteria are defined by the IEC standard 60834 as follows:

- Transmission time (Speed): The time between the moment where state changes at the transmitter input and the moment of the corresponding change at the receiver output, including propagation delay. Overall operating time for a teleprotection system includes the time for initiating the command at the transmitting end, the propagation delay over the network (including equipments) and the selection and decision time at the receiving end, including any additional delay due to a noisy environment.

- Dependability: The ability to issue and receive valid commands in the presence of interference and/or noise, by minimizing the probability of missing command (PMC). Dependability targets are typically set for a specific bit error rate (BER) level.

- Security: The ability to prevent false tripping due to a noisy environment, by minimizing the probability of unwanted commands (PUC). Security targets are also set for a specific bit error rate (BER) level.

Additional elements of the the teleprotection system that impact its performance include:

- Network bandwidth

- Failure recovery capacity (aka resiliency)

#### 3.1.1.1.2. Fault Detection and Clearance Timing

Most power line equipment can tolerate short circuits or faults for up to approximately five power cycles before sustaining irreversible damage or affecting other segments in the network. This translates to total fault clearance time of 100ms. As a safety precaution, however, actual operation time of protection systems is limited to 70- 80 percent of this period, including fault recognition time, command transmission time and line breaker switching time.

Some system components, such as large electromechanical switches, require particularly long time to operate and take up the majority of the total clearance time, leaving only a 10ms window for the telecommunications part of the protection scheme, independent of the distance to travel. Given the sensitivity of the issue, new networks impose requirements that are even more stringent: IEC standard 61850

limits the transfer time for protection messages to 1/4 - 1/2 cycle or 4 - 8ms (for 60Hz lines) for the most critical messages.

#### 3.1.1.1.3. Symmetric Channel Delay

Note: It is currently under WG discussion whether symmetric path delays are to be guaranteed by DetNet.

Teleprotection channels which are differential must be synchronous, which means that any delays on the transmit and receive paths must match each other. Teleprotection systems ideally support zero asymmetric delay; typical legacy relays can tolerate delay discrepancies of up to 750us.

Some tools available for lowering delay variation below this threshold are:

- For legacy systems using Time Division Multiplexing (TDM), jitter buffers at the multiplexers on each end of the line can be used to offset delay variation by queuing sent and received packets. The length of the queues must balance the need to regulate the rate of transmission with the need to limit overall delay, as larger buffers result in increased latency.

- For jitter-prone IP packet networks, traffic management tools can ensure that the teleprotection signals receive the highest transmission priority to minimize jitter.

- Standard packet-based synchronization technologies, such as 1588-2008 Precision Time Protocol (PTP) and Synchronous Ethernet (Sync-E), can help keep networks stable by maintaining a highly accurate clock source on the various network devices.

#### 3.1.1.1.4. Teleprotection Network Requirements (IEC 61850)

The following table captures the main network metrics as based on the IEC 61850 standard.

| Teleprotection Requirement | Attribute |
|---|---|
| One way maximum delay | 4-10 ms |
| Asymetric delay required | Yes |
| Maximum jitter | less than 250 us (750 us for legacy IED) |
| Topology | Point to point, point to Multi-point |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 0.1% to 1% |

Table 1: Teleprotection network requirements

3.1.1.1.5. Inter-Trip Protection scheme

"Inter-tripping" is the signal-controlled tripping of a circuit breaker to complete the isolation of a circuit or piece of apparatus in concert with the tripping of other circuit breakers.

| Inter-Trip protection Requirement | Attribute |
|---|---|
| One way maximum delay | 5 ms |
| Asymetric delay required | No |
| Maximum jitter | Not critical |
| Topology | Point to point, point to Multi-point |
| Bandwidth | 64 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 0.1% |

Table 2: Inter-Trip protection network requirements

3.1.1.1.6. Current Differential Protection Scheme

Current differential protection is commonly used for line protection, and is typical for protecting parallel circuits. At both end of the lines the current is measured by the differential relays, and both relays will trip the circuit breaker if the current going into the line does not equal the current going out of the line. This type of protection scheme assumes some form of communications being present between the relays at both end of the line, to allow both relays to compare measured current values. Line differential protection schemes assume a very low telecommunications delay between both relays, often as low as 5ms. Moreover, as those systems are often not time-synchronized, they also assume symmetric telecommunications paths with constant delay, which allows comparing current measurement values taken at the exact same time.

| Current Differential protection Requirement | Attribute |
|---|---|
| One way maximum delay | 5 ms |
| Asymetric delay Required | Yes |
| Maximum jitter | less than 250 us (750us for legacy IED) |
| Topology | Point to point, point to Multi-point |
| Bandwidth | 64 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 0.1% |

Table 3: Current Differential Protection metrics

3.1.1.1.7. Distance Protection Scheme

Distance (Impedance Relay) protection scheme is based on voltage and current measurements. The network metrics are similar (but not identical to) Current Differential protection.

| Distance protection Requirement | Attribute |
|---|---|
| One way maximum delay | 5 ms |
| Asymetric delay Required | No |
| Maximum jitter | Not critical |
| Topology | Point to point, point to Multi-point |
| Bandwidth | 64 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 0.1% |

Table 4: Distance Protection requirements

3.1.1.1.8. Inter-Substation Protection Signaling

This use case describes the exchange of Sampled Value and/or GOOSE (Generic Object Oriented Substation Events) message between Intelligent Electronic Devices (IED) in two substations for protection and tripping coordination. The two IEDs are in a master-slave mode.

The Current Transformer or Voltage Transformer (CT/VT) in one substation sends the sampled analog voltage or current value to the Merging Unit (MU) over hard wire. The MU sends the time-synchronized 61850-9-2 sampled values to the slave IED. The slave IED forwards the information to the Master IED in the other substation. The master IED makes the determination (for example based on sampled value differentials) to send a trip command to the originating IED. Once the slave IED/Relay receives the GOOSE trip for breaker tripping, it opens the breaker. It then sends a confirmation message back to the master. All data exchanges between IEDs are either through Sampled Value and/or GOOSE messages.

| Inter-Substation protection Requirement | Attribute |
|---|---|
| One way maximum delay | 5 ms |
| Asymetric delay Required | No |
| Maximum jitter | Not critical |
| Topology | Point to point, point to Multi-point |
| Bandwidth | 64 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 1% |

Table 5: Inter-Substation Protection requirements

#### 3.1.1.2. Intra-Substation Process Bus Communications

This use case describes the data flow from the CT/VT to the IEDs in the substation via the MU. The CT/VT in the substation send the sampled value (analog voltage or current) to the MU over hard wire. The MU sends the time-synchronized 61850-9-2 sampled values to the IEDs in the substation in GOOSE message format. The GPS Master Clock can send 1PPS or IRIG-B format to the MU through a serial port or IEEE 1588 protocol via a network. Process bus communication using 61850 simplifies connectivity within the substation and removes the requirement for multiple serial connections and removes the slow serial bus architectures that are typically used. This also ensures increased flexibility and increased speed with the use of multicast messaging between multiple devices.

| Intra-Substation protection Requirement | Attribute |
|---|---|
| One way maximum delay | 5 ms |
| Asymetric delay Required | No |
| Maximum jitter | Not critical |
| Topology | Point to point, point to Multi-point |
| Bandwidth | 64 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on Node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes - No |
| Packet loss | 0.1% |

Table 6: Intra-Substation Protection requirements

3.1.1.3. Wide Area Monitoring and Control Systems

The application of synchrophasor measurement data from Phasor Measurement Units (PMU) to Wide Area Monitoring and Control Systems promises to provide important new capabilities for improving system stability. Access to PMU data enables more timely situational awareness over larger portions of the grid than what has been possible historically with normal SCADA (Supervisory Control and Data Acquisition) data. Handling the volume and real-time nature of synchrophasor data presents unique challenges for existing application architectures. Wide Area management System (WAMS) makes it possible for the condition of the bulk power system to be observed and understood in real-time so that protective, preventative, or corrective action can be taken. Because of the very high sampling rate of measurements and the strict requirement for time synchronization of the samples, WAMS has stringent telecommunications requirements in an IP network that are captured in the following table:

| WAMS Requirement | Attribute |
|---|---|
| One way maximum delay | 50 ms |
| Asymetric delay Required | No |
| Maximum jitter | Not critical |
| Topology | Point to point, point to Multi-point, Multi-point to Multi-point |
| Bandwidth | 100 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on Node failure | less than 50ms - hitless |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 1% |

Table 7: WAMS Special Communication Requirements

#### 3.1.1.4. IEC 61850 WAN engineering guidelines requirement classification

The IEC (International Electrotechnical Commission) has recently published a Technical Report which offers guidelines on how to define and deploy Wide Area Networks for the interconnections of electric substations, generation plants and SCADA operation centers. The IEC 61850-90-12 is providing a classification of WAN communication requirements into 4 classes. Table 8 summarizes these requirements:

| WAN Requirement | Class WA | Class WB | Class WC | Class WD |
|---|---|---|---|---|
| Application field | EHV (Extra High Voltage) | HV (High Voltage) | MV (Medium Voltage) | General purpose |
| Latency | 5 ms | 10 ms | 100 ms | > 100 ms |
| Jitter | 10 us | 100 us | 1 ms | 10 ms |
| Latency Asymetry | 100 us | 1 ms | 10 ms | 100 ms |
| Time Accuracy | 1 us | 10 us | 100 us | 10 to 100 ms |
| Bit Error rate | 10-7 to 10-6 | 10-5 to 10-4 | 10-3 | |
| Unavailability | 10-7 to 10-6 | 10-5 to 10-4 | 10-3 | |
| Recovery delay | Zero | 50 ms | 5 s | 50 s |
| Cyber security | extremely high | High | Medium | Medium |

Table 8: 61850-90-12 Communication Requirements; Courtesy of IEC

### 3.1.2. Generation Use Case

The electrical power generation frequency should be maintained within a very narrow band. Deviations from the acceptable frequency range are detected and the required signals are sent to the power plants for frequency regulation.

Automatic generation control (AGC) is a system for adjusting the power output of generators at different power plants, in response to changes in the load.

| FCAG (Frequency Control Automatic Generation) Requirement | Attribute |
|---|---|
| One way maximum delay | 500 ms |
| Asymetric delay Required | No |
| Maximum jitter | Not critical |
| Topology | Point to point |
| Bandwidth | 20 Kbps |
| Availability | 99.999 |
| precise timing required | Yes |
| Recovery time on Node failure | N/A |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 1% |

Table 9: FCAG Communication Requirements

### 3.1.3. Distribution use case

#### 3.1.3.1. Fault Location Isolation and Service Restoration (FLISR)

Fault Location, Isolation, and Service Restoration (FLISR) refers to the ability to automatically locate the fault, isolate the fault, and restore service in the distribution network. This will likely be the first widespread application of distributed intelligence in the grid.

Static power switch status (open/closed) in the network dictates the power flow to secondary substations. Reconfiguring the network in the event of a fault is typically done manually on site to energize/ de-energize alternate paths. Automating the operation of substation switchgear allows the flow of power to be altered automatically under fault conditions.

FLISR can be managed centrally from a Distribution Management System (DMS) or executed locally through distributed control via intelligent switches and fault sensors.

| FLISR Requirement | Attribute |
|---|---|
| One way maximum delay | 80 ms |
| Asymetric delay Required | No |
| Maximum jitter | 40 ms |
| Topology | Point to point, point to Multi-point, Multi-point to Multi-point |
| Bandwidth | 64 Kbps |
| Availability | 99.9999 |
| precise timing required | Yes |
| Recovery time on Node failure | Depends on customer impact |
| performance management | Yes, Mandatory |
| Redundancy | Yes |
| Packet loss | 0.1% |

Table 10: FLISR Communication Requirements

## 3.2. Electrical Utilities Today

Many utilities still rely on complex environments formed of multiple application-specific proprietary networks, including TDM networks.

In this kind of environment there is no mixing of OT and IT applications on the same network, and information is siloed between operational areas.

Specific calibration of the full chain is required, which is costly.

This kind of environment prevents utility operations from realizing the operational efficiency benefits, visibility, and functional integration of operational information across grid applications and data networks.

In addition, there are many security-related issues as discussed in the following section.

### 3.2.1. Security Current Practices and Limitations

Grid monitoring and control devices are already targets for cyber attacks, and legacy telecommunications protocols have many intrinsic network-related vulnerabilities. For example, DNP3, Modbus,

PROFIBUS/PROFINET, and other protocols are designed around a common paradigm of request and respond. Each protocol is designed for a master device such as an HMI (Human Machine Interface) system to send commands to subordinate slave devices to retrieve data (reading inputs) or control (writing to outputs). Because many of these protocols lack authentication, encryption, or other basic security measures, they are prone to network-based attacks, allowing a malicious actor or attacker to utilize the request-and-respond system as a mechanism for command-and-control like functionality. Specific security concerns common to most industrial control, including utility telecommunication protocols include the following:

- Network or transport errors (e.g. malformed packets or excessive latency) can cause protocol failure.
- Protocol commands may be available that are capable of forcing slave devices into inoperable states, including powering-off devices, forcing them into a listen-only state, disabling alarming.
- Protocol commands may be available that are capable of restarting communications and otherwise interrupting processes.
- Protocol commands may be available that are capable of clearing, erasing, or resetting diagnostic information such as counters and diagnostic registers.
- Protocol commands may be available that are capable of requesting sensitive information about the controllers, their configurations, or other need-to-know information.
- Most protocols are application layer protocols transported over TCP; therefore it is easy to transport commands over non-standard ports or inject commands into authorized traffic flows.
- Protocol commands may be available that are capable of broadcasting messages to many devices at once (i.e. a potential DoS).
- Protocol commands may be available to query the device network to obtain defined points and their values (i.e. a configuration scan).
- Protocol commands may be available that will list all available function codes (i.e. a function scan).

These inherent vulnerabilities, along with increasing connectivity between IT an OT networks, make network-based attacks very feasible.

Simple injection of malicious protocol commands provides control over the target process. Altering legitimate protocol traffic can also alter information about a process and disrupt the legitimate controls that are in place over that process. A man-in-the-middle attack could provide both control over a process and misrepresentation of data back to operator consoles.

### 3.3. Electrical Utilities Future

The business and technology trends that are sweeping the utility industry will drastically transform the utility business from the way it has been for many decades. At the core of many of these changes is a drive to modernize the electrical grid with an integrated telecommunications infrastructure. However, interoperability concerns, legacy networks, disparate tools, and stringent security requirements all add complexity to the grid transformation. Given the range and diversity of the requirements that should be addressed by the next generation telecommunications infrastructure, utilities need to adopt a holistic architectural approach to integrate the electrical grid with digital telecommunications across the entire power delivery chain.

The key to modernizing grid telecommunications is to provide a common, adaptable, multi-service network infrastructure for the entire utility organization. Such a network serves as the platform for current capabilities while enabling future expansion of the network to accommodate new applications and services.

To meet this diverse set of requirements, both today and in the future, the next generation utility telecommunnications network will be based on open-standards-based IP architecture. An end-to-end IP architecture takes advantage of nearly three decades of IP technology development, facilitating interoperability across disparate networks and devices, as it has been already demonstrated in many mission-critical and highly secure networks.

IPv6 is seen as a future telecommunications technology for the Smart Grid; the IEC (International Electrotechnical Commission) and different National Committees have mandated a specific adhoc group (AHG8) to define the migration strategy to IPv6 for all the IEC TC57 power automation standards.

#### 3.3.1. Migration to Packet-Switched Network

Throughout the world, utilities are increasingly planning for a future based on smart grid applications requiring advanced telecommunications systems. Many of these applications utilize packet connectivity for communicating information and control signals

across the utility's Wide Area Network (WAN), made possible by technologies such as multiprotocol label switching (MPLS). The data that traverses the utility WAN includes:

- Grid monitoring, control, and protection data
- Non-control grid data (e.g. asset data for condition-based monitoring)
- Physical safety and security data (e.g. voice and video)
- Remote worker access to corporate applications (voice, maps, schematics, etc.)
- Field area network backhaul for smart metering, and distribution grid management
- Enterprise traffic (email, collaboration tools, business applications)

WANs support this wide variety of traffic to and from substations, the transmission and distribution grid, generation sites, between control centers, and between work locations and data centers. To maintain this rapidly expanding set of applications, many utilities are taking steps to evolve present time-division multiplexing (TDM) based and frame relay infrastructures to packet systems. Packet-based networks are designed to provide greater functionalities and higher levels of service for applications, while continuing to deliver reliability and deterministic (real-time) traffic support.

### 3.3.2. Telecommunications Trends

These general telecommunications topics are in addition to the use cases that have been addressed so far. These include both current and future telecommunications related topics that should be factored into the network architecture and design.

#### 3.3.2.1. General Telecommunications Requirements

- IP Connectivity everywhere
- Monitoring services everywhere and from different remote centers
- Move services to a virtual data center
- Unify access to applications / information from the corporate network

- Unify services
- Unified Communications Solutions
- Mix of fiber and microwave technologies - obsolescence of SONET/ SDH or TDM
- Standardize grid telecommunications protocol to opened standard to ensure interoperability
- Reliable Telecommunications for Transmission and Distribution Substations
- IEEE 1588 time synchronization Client / Server Capabilities
- Integration of Multicast Design
- QoS Requirements Mapping
- Enable Future Network Expansion
- Substation Network Resilience
- Fast Convergence Design
- Scalable Headend Design
- Define Service Level Agreements (SLA) and Enable SLA Monitoring
- Integration of 3G/4G Technologies and future technologies
- Ethernet Connectivity for Station Bus Architecture
- Ethernet Connectivity for Process Bus Architecture
- Protection, teleprotection and PMU (Phaser Measurement Unit) on IP

#### 3.3.2.2. Specific Network topologies of Smart Grid Applications

Utilities often have very large private telecommunications networks. It covers an entire territory / country. The main purpose of the network, until now, has been to support transmission network monitoring, control, and automation, remote control of generation sites, and providing FCAPS (Fault, Configuration, Accounting, Performance, Security) services from centralized network operation centers.

Going forward, one network will support operation and maintenance of electrical networks (generation, transmission, and distribution), voice and data services for ten of thousands of employees and for exchange with neighboring interconnections, and administrative services. To meet those requirements, utility may deploy several physical networks leveraging different technologies across the country: an optical network and a microwave network for instance. Each protection and automatism system between two points has two telecommunications circuits, one on each network. Path diversity between two substations is key. Regardless of the event type (hurricane, ice storm, etc.), one path shall stay available so the system can still operate.

In the optical network, signals are transmitted over more than tens of thousands of circuits using fiber optic links, microwave and telephone cables. This network is the nervous system of the utility's power transmission operations. The optical network represents ten of thousands of km of cable deployed along the power lines, with individual runs as long as 280 km.

#### 3.3.2.3. Precision Time Protocol

Some utilities do not use GPS clocks in generation substations. One of the main reasons is that some of the generation plants are 30 to 50 meters deep under ground and the GPS signal can be weak and unreliable. Instead, atomic clocks are used. Clocks are synchronized amongst each other. Rubidium clocks provide clock and 1ms timestamps for IRIG-B.

Some companies plan to transition to the Precision Time Protocol (PTP, [IEEE1588]), distributing the synchronization signal over the IP/MPLS network. PTP provides a mechanism for synchronizing the clocks of participating nodes to a high degree of accuracy and precision.

PTP operates based on the following assumptions:

It is assumed that the network eliminates cyclic forwarding of PTP messages within each communication path (e.g. by using a spanning tree protocol).

PTP is tolerant of an occasional missed message, duplicated message, or message that arrived out of order. However, PTP assumes that such impairments are relatively rare.

PTP was designed assuming a multicast communication model, however PTP also supports a unicast communication model as long as the behavior of the protocol is preserved.

Like all message-based time transfer protocols, PTP time accuracy is degraded by delay asymmetry in the paths taken by event messages. Asymmetry is not detectable by PTP, however, if such delays are known a priori, PTP can correct for asymmetry.

IEC 61850 will recommend the use of the IEEE PTP 1588 Utility Profile (as defined in [IEC62439-3:2012] Annex B) which offers the support of redundant attachment of clocks to Parallel Redundancy Protcol (PRP) and High-availability Seamless Redundancy (HSR) networks.

### 3.3.3. Security Trends in Utility Networks

Although advanced telecommunications networks can assist in transforming the energy industry by playing a critical role in maintaining high levels of reliability, performance, and manageability, they also introduce the need for an integrated security infrastructure. Many of the technologies being deployed to support smart grid projects such as smart meters and sensors can increase the vulnerability of the grid to attack. Top security concerns for utilities migrating to an intelligent smart grid telecommunications platform center on the following trends:

- Integration of distributed energy resources
- Proliferation of digital devices to enable management, automation, protection, and control
- Regulatory mandates to comply with standards for critical infrastructure protection
- Migration to new systems for outage management, distribution automation, condition-based maintenance, load forecasting, and smart metering
- Demand for new levels of customer service and energy management

This development of a diverse set of networks to support the integration of microgrids, open-access energy competition, and the use of network-controlled devices is driving the need for a converged security infrastructure for all participants in the smart grid, including utilities, energy service providers, large commercial and industrial, as well as residential customers. Securing the assets of electric power delivery systems (from the control center to the substation, to the feeders and down to customer meters) requires an end-to-end security infrastructure that protects the myriad of telecommunications assets used to operate, monitor, and control power flow and measurement.

"Cyber security" refers to all the security issues in automation and telecommunications that affect any functions related to the operation of the electric power systems. Specifically, it involves the concepts of:

- Integrity : data cannot be altered undetectably
- Authenticity : the telecommunications parties involved must be validated as genuine
- Authorization : only requests and commands from the authorized users can be accepted by the system
- Confidentiality : data must not be accessible to any unauthenticated users

When designing and deploying new smart grid devices and telecommunications systems, it is imperative to understand the various impacts of these new components under a variety of attack situations on the power grid. Consequences of a cyber attack on the grid telecommunications network can be catastrophic. This is why security for smart grid is not just an ad hoc feature or product, it's a complete framework integrating both physical and Cyber security requirements and covering the entire smart grid networks from generation to distribution. Security has therefore become one of the main foundations of the utility telecom network architecture and must be considered at every layer with a defense-in-depth approach. Migrating to IP based protocols is key to address these challenges for two reasons:

- IP enables a rich set of features and capabilities to enhance the security posture
- IP is based on open standards, which allows interoperability between different vendors and products, driving down the costs associated with implementing security solutions in OT networks.

Securing OT (Operation technology) telecommunications over packet-switched IP networks follow the same principles that are foundational for securing the IT infrastructure, i.e., consideration must be given to enforcing electronic access control for both person-to-machine and machine-to-machine communications, and providing the appropriate levels of data privacy, device and platform integrity, and threat detection and mitigation.

### 3.4. Electrical Utilities Asks

- Mixed L2 and L3 topologies
- Deterministic behavior
- Bounded latency and jitter
- High availability, low recovery time
- Redundancy, low packet loss
- Precise timing
- Centralized computing of deterministic paths
- Distributed configuration may also be useful

## 4. Building Automation Systems

### 4.1. Use Case Description

A Building Automation System (BAS) manages equipment and sensors in a building for improving residents' comfort, reducing energy consumption, and responding to failures and emergencies. For example, the BAS measures the temperature of a room using sensors and then controls the HVAC (heating, ventilating, and air conditioning) to maintain a set temperature and minimize energy consumption.

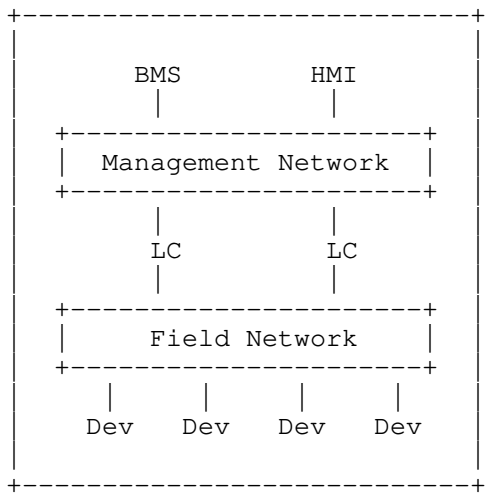
A BAS primarily performs the following functions:

- Periodically measures states of devices, for example humidity and illuminance of rooms, open/close state of doors, FAN speed, etc.
- Stores the measured data.
- Provides the measured data to BAS systems and operators.
- Generates alarms for abnormal state of devices.
- Controls devices (e.g. turn off room lights at 10:00 PM).

### 4.2. Building Automation Systems Today

4.2.1. BAS Architecture

A typical BAS architecture of today is shown in Figure 1.




BMS := Building Management Server
HMI := Human Machine Interface
LC := Local Controller

Figure 1: BAS architecture

There are typically two layers of network in a BAS. The upper one is called the Management Network and the lower one is called the Field Network. In management networks an IP-based communication protocol is used, while in field networks non-IP based communication protocols ("field protocols") are mainly used. Field networks have specific timing requirements, whereas management networks can be best-effort.

A Human Machine Interface (HMI) is typically a desktop PC used by operators to monitor and display device states, send device control commands to Local Controllers (LCs), and configure building schedules (for example "turn off all room lights in the building at 10:00 PM").

A Building Management Server (BMS) performs the following operations.

- o Collect and store device states from LCs at regular intervals.
- o Send control values to LCs according to a building schedule.
- o Send an alarm signal to operators if it detects abnormal devices states.

The BMS and HMI communicate with LCs via IP-based "management protocols" (see standards [bacnetip], [knx]).

A LC is typically a Programmable Logic Controller (PLC) which is connected to several tens or hundreds of devices using "field protocols". An LC performs the following kinds of operations:

- Measure device states and provide the information to BMS or HMI.
- Send control values to devices, unilaterally or as part of a feedback control loop.

There are many field protocols used today; some are standards-based and others are proprietary (see standards [lontalk], [modbus], [profibus] and [flnet]). The result is that BASs have multiple MAC/PHY modules and interfaces. This makes BASs more expensive, slower to develop, and can result in "vendor lock-in" with multiple types of management applications.

#### 4.2.2. BAS Deployment Model

An example BAS for medium or large buildings is shown in Figure 2. The physical layout spans multiple floors, and there is a monitoring room where the BAS management entities are located. Each floor will have one or more LCs depending upon the number of devices connected to the field network.

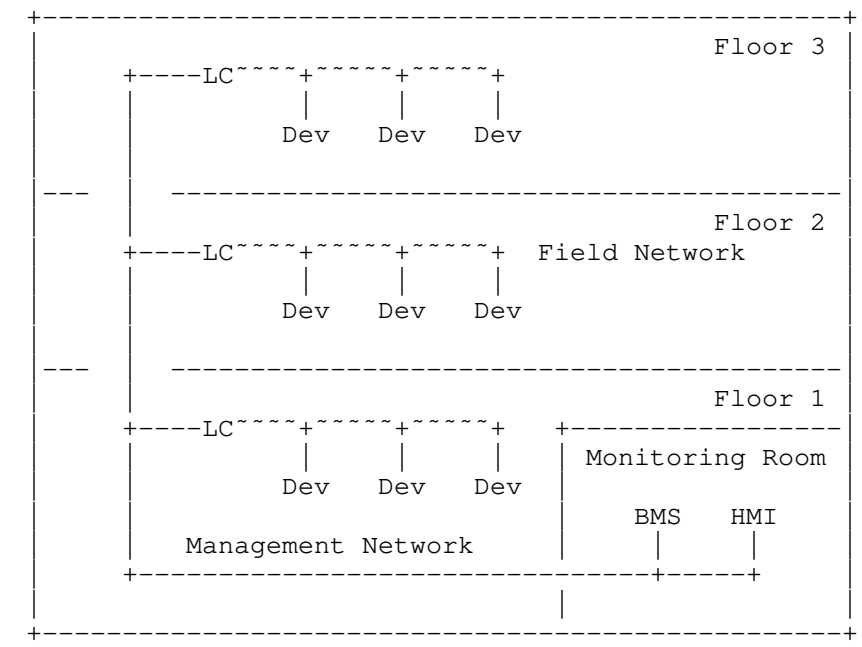



Figure 2: BAS Deployment model for Medium/Large Buildings

Each LC is connected to the monitoring room via the Management network, and the management functions are performed within the building. In most cases, fast Ethernet (e.g. 100BASE-T) is used for the management network. Since the management network is non-realtime, use of Ethernet without quality of service is sufficient for today's deployment.

In the field network a variety of physical interfaces such as RS232C and RS485 are used, which have specific timing requirements. Thus if a field network is to be replaced with an Ethernet or wireless network, such networks must support time-critical deterministic flows.

In Figure 3, another deployment model is presented in which the management system is hosted remotely. This is becoming popular for small office and residential buildings in which a standalone monitoring system is not cost-effective.

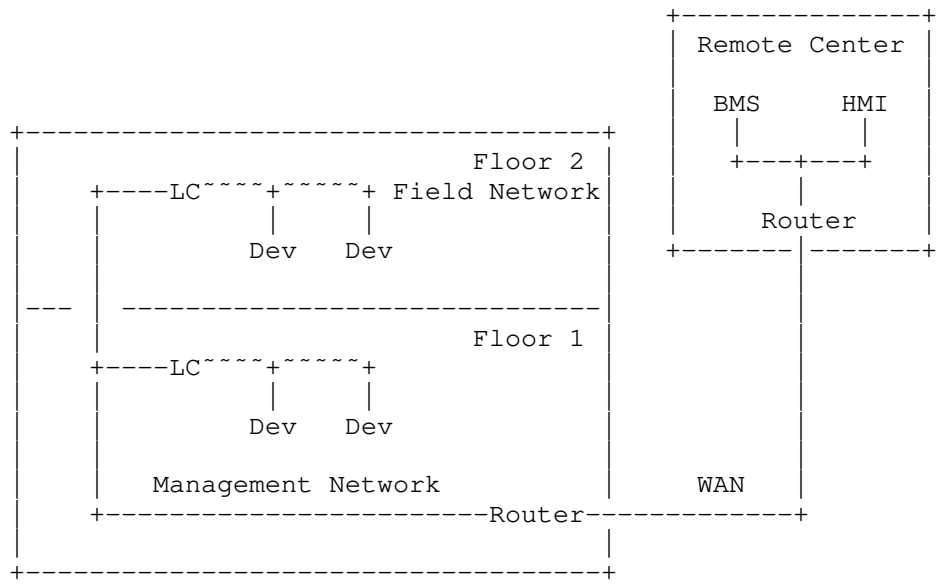



Figure 3: Deployment model for Small Buildings

Some interoperability is possible today in the Management Network, but not in today's field networks due to their non-IP-based design.

#### 4.2.3. Use Cases for Field Networks

Below are use cases for Environmental Monitoring, Fire Detection, and Feedback Control, and their implications for field network performance.

##### 4.2.3.1. Environmental Monitoring

The BMS polls each LC at a maximum measurement interval of 100ms (for example to draw a historical chart of 1 second granularity with a 10x sampling interval) and then performs the operations as specified by the operator. Each LC needs to measure each of its several hundred sensors once per measurement interval. Latency is not critical in this scenario as long as all sensor values are completed in the measurement interval. Availability is expected to be 99.999 %.

##### 4.2.3.2. Fire Detection

On detection of a fire, the BMS must stop the HVAC, close the fire shutters, turn on the fire sprinklers, send an alarm, etc. There are typically ~10s of sensors per LC that BMS needs to manage. In this

scenario the measurement interval is 10-50ms, the communication delay is 10ms, and the availability must be 99.9999 %.

#### 4.2.3.3. Feedback Control

BAS systems utilize feedback control in various ways; the most time-critial is control of DC motors, which require a short feedback interval (1-5ms) with low communication delay (10ms) and jitter (1ms). The feedback interval depends on the characteristics of the device and a target quality of control value. There are typically ~10s of such devices per LC.

Communication delay is expected to be less than 10 ms, jitter less than 1 sec while the availability must be 99.9999% .

### 4.2.4. Security Considerations

When BAS field networks were developed it was assumed that the field networks would always be physically isolated from external networks and therefore security was not a concern. In today's world many BASs are managed remotely and are thus connected to shared IP networks and so security is definitely a concern, yet security features are not available in the majority of BAS field network deployments .

The management network, being an IP-based network, has the protocols available to enable network security, but in practice many BAS systems do not implement even the available security features such as device authentication or encryption for data in transit.

## 4.3. BAS Future

In the future we expect more fine-grained environmental monitoring and lower energy consumption, which will require more sensors and devices, thus requiring larger and more complex building networks.

We expect building networks to be connected to or converged with other networks (Enterprise network, Home network, and Internet).

Therefore better facilities for network management, control, reliability and security are critical in order to improve resident and operator convenience and comfort. For example the ability to monitor and control building devices via the internet would enable (for example) control of room lights or HVAC from a resident's desktop PC or phone application.

### 4.4. BAS Asks

The community would like to see an interoperable protocol specification that can satisfy the timing, security, availability and QoS constraints described above, such that the resulting converged network can replace the disparate field networks. Ideally this connectivity could extend to the open Internet.

This would imply an architecture that can guarantee

- Low communication delays (from <10ms to 100ms in a network of several hundred devices)
- Low jitter (< 1 ms)
- Tight feedback intervals (1ms – 10ms)
- High network availability (up to 99.9999% )
- Availability of network data in disaster scenario
- Authentication between management and field devices (both local and remote)
- Integrity and data origin authentication of communication data between field and management devices
- Confidentiality of data when communicated to a remote device

## 5. Wireless for Industrial

### 5.1. Use Case Description

Wireless networks are useful for industrial applications, for example when portable, fast-moving or rotating objects are involved, and for the resource-constrained devices found in the Internet of Things (IoT).

Such network-connected sensors, actuators, control loops (etc.) typically require that the underlying network support real-time quality of service (QoS), as well as specific classes of other network properties such as reliability, redundancy, and security.

These networks may also contain very large numbers of devices, for example for factories, "big data" acquisition, and the IoT. Given the large numbers of devices installed, and the potential pervasiveness of the IoT, this is a huge and very cost-sensitive

market. For example, a 1% cost reduction in some areas could save $100B

#### 5.1.1. Network Convergence using 6TiSCH

Some wireless network technologies support real-time QoS, and are thus useful for these kinds of networks, but others do not. For example WiFi is pervasive but does not provide guaranteed timing or delivery of packets, and thus is not useful in this context.

In this use case we focus on one specific wireless network technology which does provide the required deterministic QoS, which is "IPv6 over the TSCH mode of IEEE 802.15.4e" (6TiSCH, where TSCH stands for "Time-Slotted Channel Hopping", see [I-D.ietf-6tisch-architecture], [IEEE802154], [IEEE802154e], and [RFC7554]).

There are other deterministic wireless busses and networks available today, however they are imcompatible with each other, and incompatible with IP traffic (for example [ISA100], [WirelessHART]).

Thus the primary goal of this use case is to apply 6TiSH as a converged IP- and standards-based wireless network for industrial applications, i.e. to replace multiple proprietary and/or incompatible wireless networking and wireless network management standards.

#### 5.1.2. Common Protocol Development for 6TiSCH

Today there are a number of protocols required by 6TiSCH which are still in development, and a second intent of this use case is to highlight the ways in which these "missing" protocols share goals in common with DetNet. Thus it is possible that some of the protocol technology developed for DetNet will also be applicable to 6TiSCH.

These protocol goals are identified here, along with their relationship to DetNet. It is likely that ultimately the resulting protocols will not be identical, but will share design principles which contribute to the eficiency of enabling both DetNet and 6TiSCH.

One such commonality is that although at a different time scale, in both TSN [IEEE802.1TSNTG] and TSCH a packet crosses the network from node to node follows a precise schedule, as a train that leaves intermediate stations at precise times along its path. This kind of operation reduces collisions, saves energy, and enables engineering the network for deterministic properties.

Another commonality is remote monitoring and scheduling management of a TSCH network by a Path Computation Element (PCE) and Network

Management Entity (NME). The PCE/NME manage timeslots and device resources in a manner that minimizes the interaction with and the load placed on resource-constrained devices. For example, a tiny IoT device may have just enough buffers to store one or a few IPv6 packets, and will have limited bandwidth between peers such that it can maintain only a small amount of peer information, and will not be able to store many packets waiting to be forwarded. It is advantageous then for it to only be required to carry out the specific behavior assigned to it by the PCE/NME (as opposed to maintaining its own IP stack, for example).

Note: Current WG discussion indicates that some peer-to-peer communication must be assumed, i.e. the PCE may communicate only indirectly with any given device, enabling hierarchical configuration of the system.

6TiSCH depends on [PCE] and [I-D.finn-detnet-architecture].

6TiSCH also depends on the fact that DetNet will maintain consistency with [IEEE802.1TSNTG].

## 5.2. Wireless Industrial Today

Today industrial wireless is accomplished using multiple deterministic wireless networks which are incompatible with each other and with IP traffic.

6TiSCH is not yet fully specified, so it cannot be used in today's applications.

## 5.3. Wireless Industrial Future

### 5.3.1. Unified Wireless Network and Management

We expect DetNet and 6TiSCH together to enable converged transport of deterministic and best-effort traffic flows between real-time industrial devices and wide area networks via IP routing. A high level view of a basic such network is shown in Figure 4.

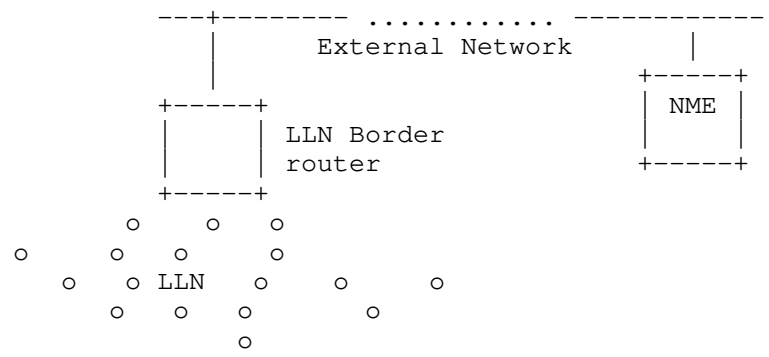



Figure 4: Basic 6TiSCH Network

Figure 5 shows a backbone router federating multiple synchronized 6TiSCH subnets into a single subnet connected to the external network.

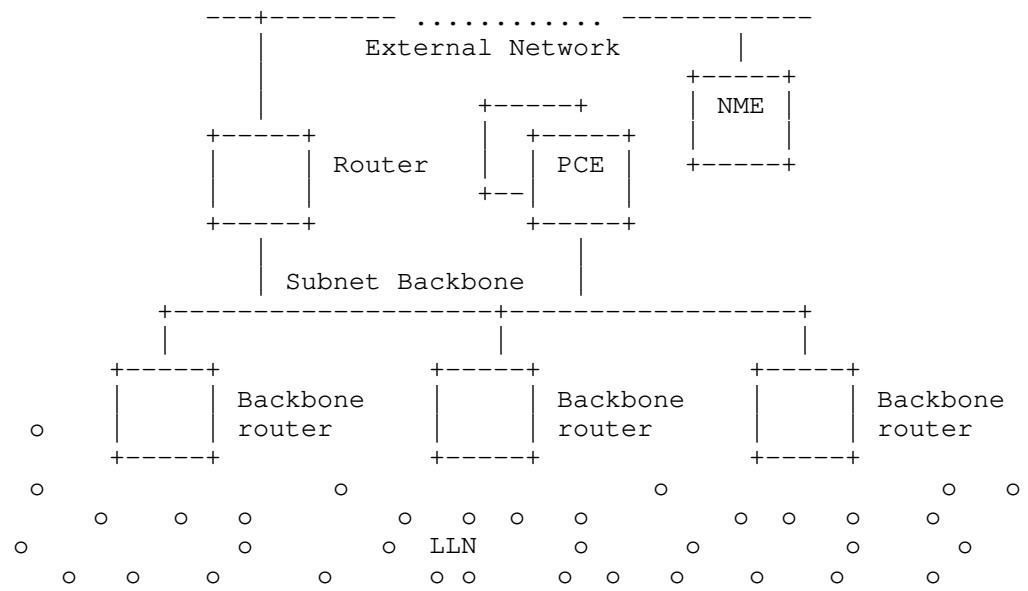



Figure 5: Extended 6TiSCH Network

The backbone router must ensure end-to-end deterministic behavior between the LLN and the backbone. We would like to see this accomplished in conformance with the work done in [I-D.finn-detnet-architecture] with respect to Layer-3 aspects of deterministic networks that span multiple Layer-2 domains.

The PCE must compute a deterministic path end-to-end across the TSCH network and IEEE802.1 TSN Ethernet backbone, and DetNet protocols are expected to enable end-to-end deterministic forwarding.

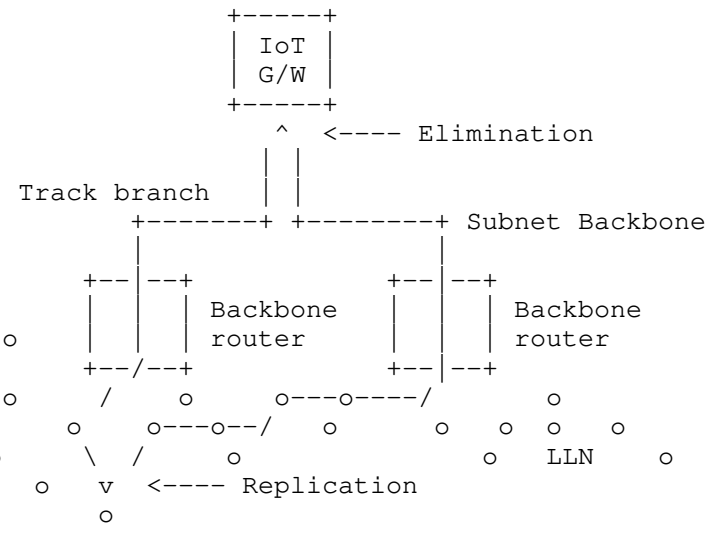



Figure 6: 6TiSCH Network with PRE

5.3.1.1. PCE and 6TiSCH ARQ Retries

Note: The possible use of ARQ techniques in DetNet is currently considered a possible design alternative.

6TiSCH uses the IEEE802.15.4 Automatic Repeat-reQuest (ARQ) mechanism to provide higher reliability of packet delivery. ARQ is related to packet replication and elimination because there are two independent paths for packets to arrive at the destination, and if an expected packed does not arrive on one path then it checks for the packet on the second path.

Although to date this mechanism is only used by wireless networks, this may be a technique that would be appropriate for DetNet and so aspects of the enabling protocol could be co-developed.

For example, in Figure 6, a Track is laid out from a field device in a 6TiSCH network to an IoT gateway that is located on a IEEE802.1 TSN backbone.

In ARQ the Replication function in the field device sends a copy of each packet over two different branches, and the PCE schedules each hop of both branches so that the two copies arrive in due time at the gateway. In case of a loss on one branch, hopefully the other copy of the packet still arrives within the allocated time. If two copies make it to the IoT gateway, the Elimination function in the gateway ignores the extra packet and presents only one copy to upper layers.

At each 6TiSCH hop along the Track, the PCE may schedule more than one timeSlot for a packet, so as to support Layer-2 retries (ARQ).

In current deployments, a TSCH Track does not necessarily support PRE but is systematically multi-path. This means that a Track is scheduled so as to ensure that each hop has at least two forwarding solutions, and the forwarding decision is to try the preferred one and use the other in case of Layer-2 transmission failure as detected by ARQ.

#### 5.3.2. Schedule Management by a PCE

A common feature of 6TiSCH and DetNet is the action of a PCE to configure paths through the network. Specifically, what is needed is a protocol and data model that the PCE will use to get/set the relevant configuration from/to the devices, as well as perform operations on the devices. We expect that this protocol will be developed by DetNet with consideration for its reuse by 6TiSCH. The remainder of this section provides a bit more context from the 6TiSCH side.

##### 5.3.2.1. PCE Commands and 6TiSCH CoAP Requests

The 6TiSCH device does not expect to place the request for bandwidth between itself and another device in the network. Rather, an operation control system invoked through a human interface specifies the required traffic specification and the end nodes (in terms of latency and reliability). Based on this information, the PCE must compute a path between the end nodes and provision the network with per-flow state that describes the per-hop operation for a given packet, the corresponding timeslots, and the flow identification that enables recognizing that a certain packet belongs to a certain path, etc.

For a static configuration that serves a certain purpose for a long period of time, it is expected that a node will be provisioned in one shot with a full schedule, which incorporates the aggregation of its behavior for multiple paths. 6TiSCH expects that the programing of the schedule will be done over COAP as discussed in [I-D.ietf-6tisch-coap].

6TiSCH expects that the PCE commands will be mapped back and forth into CoAP by a gateway function at the edge of the 6TiSCH network. For instance, it is possible that a mapping entity on the backbone transforms a non-CoAP protocol such as PCEP into the RESTful interfaces that the 6TiSCH devices support. This architecture will be refined to comply with DetNet [I-D.finn-detnet-architecture] when the work is formalized. Related information about 6TiSCH can be found at [I-D.ietf-6tisch-6top-interface] and RPL [RFC6550].

A protocol may be used to update the state in the devices during runtime, for example if it appears that a path through the network has ceased to perform as expected, but in 6TiSCH that flow was not designed and no protocol was selected. We would like to see DetNet define the appropriate end-to-end protocols to be used in that case. The implication is that these state updates take place once the system is configured and running, i.e. they are not limited to the initial communication of the configuration of the system.

A "slotFrame" is the base object that a PCE would manipulate to program a schedule into an LLN node ([I-D.ietf-6tisch-architecture]).

We would like to see the PCE read energy data from devices, and compute paths that will implement policies on how energy in devices is consumed, for instance to ensure that the spent energy does not exceeded the available energy over a period of time. Note: this statement implies that an extensible protocol for communicating device info to the PCE and enabling the PCE to act on it will be part of the DetNet architecture, however for subnets with specific protocols (e.g. CoAP) a gateway may be required.

6TiSCH devices can discover their neighbors over the radio using a mechanism such as beacons, but even though the neighbor information is available in the 6TiSCH interface data model, 6TiSCH does not describe a protocol to proactively push the neighborhood information to a PCE. We would like to see DetNet define such a protocol; one possible design alternative is that it could operate over CoAP, alternatively it could be converted to/from CoAP by a gateway. We would like to see such a protocol carry multiple metrics, for example similar to those used for RPL operations [RFC6551]

#### 5.3.2.2. 6TiSCH IP Interface

"6top" ([I-D.wang-6tisch-6top-sublayer]) is a logical link control sitting between the IP layer and the TSCH MAC layer which provides the link abstraction that is required for IP operations. The 6top data model and management interfaces are further discussed in [I-D.ietf-6tisch-6top-interface] and [I-D.ietf-6tisch-coap].

An IP packet that is sent along a 6TiSCH path uses the Differentiated Services Per-Hop-Behavior Group called Deterministic Forwarding, as described in [I-D.svshah-tsvwg-deterministic-forwarding].

#### 5.3.3. 6TiSCH Security Considerations

On top of the classical requirements for protection of control signaling, it must be noted that 6TiSCH networks operate on limited resources that can be depleted rapidly in a DoS attack on the system, for instance by placing a rogue device in the network, or by obtaining management control and setting up unexpected additional paths.

### 5.4. Wireless Industrial Asks

6TiSCH depends on DetNet to define:

- Configuration (state) and operations for deterministic paths
- End-to-end protocols for deterministic forwarding (tagging, IP)
- Protocol for packet replication and elimination

## 6. Cellular Radio

### 6.1. Use Case Description

This use case describes the application of deterministic networking in the context of cellular telecom transport networks. Important elements include time synchronization, clock distribution, and ways of establishing time-sensitive streams for both Layer-2 and Layer-3 user plane traffic.

#### 6.1.1. Network Architecture

Figure 7 illustrates a typical 3GPP-defined cellular network architecture, which includes "Fronthaul" and "Midhaul" network segments. The "Fronthaul" is the network connecting base stations (baseband processing units) to the remote radio heads (antennas). The "Midhaul" is the network inter-connecting base stations (or small cell sites).

In Figure 7 "eNB" ("E-UTRAN Node B") is the hardware that is connected to the mobile phone network which communicates directly with mobile handsets ([TS36300]).

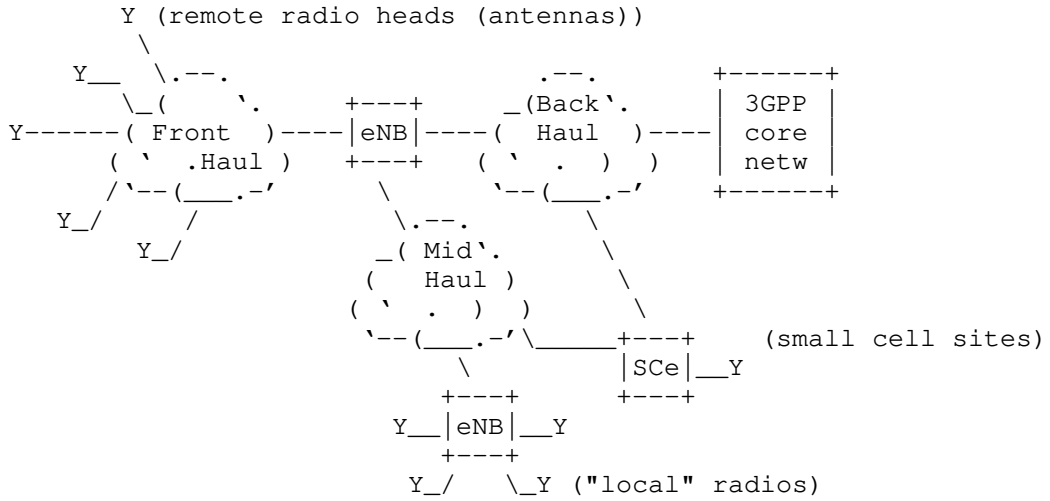



Figure 7: Generic 3GPP-based Cellular Network Architecture

### 6.1.2. Delay Constraints

The available processing time for Fronthaul networking overhead is limited to the available time after the baseband processing of the radio frame has completed. For example in Long Term Evolution (LTE) radio, processing of a radio frame is allocated 3ms but typically the processing uses most of it, allowing only a small fraction to be used by the Fronthaul network (e.g. up to 250us one-way delay, though the existing spec ([NGMN-fronth]) supports delay only up to 100us). This ultimately determines the distance the remote radio heads can be located from the base stations (e.g., 100us equals roughly 20 km of optical fiber-based transport). Allocation options of the available time budget between processing and transport are under heavy discussions in the mobile industry.

For packet-based transport the allocated transport time (e.g. CPRI would allow for 100us delay [CPRI]) is consumed by all nodes and buffering between the remote radio head and the baseband processing unit, plus the distance-incurred delay.

The baseband processing time and the available "delay budget" for the fronthaul is likely to change in the forthcoming "5G" due to reduced radio round trip times and other architectural and service requirements [NGMN].

[METIS] documents the fundamental challenges as well as overall technical goals of the future 5G mobile and wireless system as the starting point. These future systems should support much higher data

volumes and rates and significantly lower end-to-end latency for 100x more connected devices (at similar cost and energy consumption levels as today's system).

For Midhaul connections, delay constraints are driven by Inter-Site radio functions like Coordinated Multipoint Processing (CoMP, see [CoMP]). CoMP reception and transmission is a framework in which multiple geographically distributed antenna nodes cooperate to improve the performance of the users served in the common cooperation area. The design principal of CoMP is to extend the current single-cell to multi-UE (User Equipment) transmission to a multi-cell-to-multi-UEs transmission by base station cooperation.

CoMP has delay-sensitive performance parameters, which are "midhaul latency" and "CSI (Channel State Information) reporting and accuracy". The essential feature of CoMP is signaling between eNBs, so Midhaul latency is the dominating limitation of CoMP performance. Generally, CoMP can benefit from coordinated scheduling (either distributed or centralized) of different cells if the signaling delay between eNBs is within 1-10ms. This delay requirement is both rigid and absolute because any uncertainty in delay will degrade the performance significantly.

Inter-site CoMP is one of the key requirements for 5G and is also a near-term goal for the current 4.5G network architecture.

#### 6.1.3. Time Synchronization Constraints

Fronthaul time synchronization requirements are given by [TS25104], [TS36104], [TS36211], and [TS36133]. These can be summarized for the current 3GPP LTE-based networks as:

Delay Accuracy:
: +-8ns (i.e. +-1/32 Tc, where Tc is the UMTS Chip time of 1/3.84 MHz) resulting in a round trip accuracy of +-16ns. The value is this low to meet the 3GPP Timing Alignment Error (TAE) measurement requirements. Note: performance guarantees of low nanosecond values such as these are considered to be below the DetNet layer - it is assumed that the underlying implementation, e.g. the hardware, will provide sufficient support (e.g. buffering) to enable this level of accuracy. These values are maintained in the use case to give an indication of the overall application.

Timing Alignment Error:
: Timing Alignment Error (TAE) is problematic to Fronthaul networks and must be minimized. If the transport network cannot guarantee low enough TAE then additional buffering has to be introduced at the edges of the network to buffer out the jitter. Buffering is

not desirable as it reduces the total available delay budget. Packet Delay Variation (PDV) requirements can be derived from TAE for packet based Fronthaul networks.

* For multiple input multiple output (MIMO) or TX diversity transmissions, at each carrier frequency, TAE shall not exceed 65 ns (i.e. 1/4 Tc).

* For intra-band contiguous carrier aggregation, with or without MIMO or TX diversity, TAE shall not exceed 130 ns (i.e. 1/2 Tc).

* For intra-band non-contiguous carrier aggregation, with or without MIMO or TX diversity, TAE shall not exceed 260 ns (i.e. one Tc).

* For inter-band carrier aggregation, with or without MIMO or TX diversity, TAE shall not exceed 260 ns.

Transport link contribution to radio frequency error:
+-2 PPB. This value is considered to be "available" for the Fronthaul link out of the total 50 PPB budget reserved for the radio interface. Note: the reason that the transport link contributes to radio frequency error is as follows. The current way of doing Fronthaul is from the radio unit to remote radio head directly. The remote radio head is essentially a passive device (without buffering etc.) The transport drives the antenna directly by feeding it with samples and everything the transport adds will be introduced to radio as-is. So if the transport causes additional frequency error that shows immediately on the radio as well. Note: performance guarantees of low nanosecond values such as these are considered to be below the DetNet layer - it is assumed that the underlying implementation, e.g. the hardware, will provide sufficient support to enable this level of performance. These values are maintained in the use case to give an indication of the overall application.

The above listed time synchronization requirements are difficult to meet with point-to-point connected networks, and more difficult when the network includes multiple hops. It is expected that networks must include buffering at the ends of the connections as imposed by the jitter requirements, since trying to meet the jitter requirements in every intermediate node is likely to be too costly. However, every measure to reduce jitter and delay on the path makes it easier to meet the end-to-end requirements.

In order to meet the timing requirements both senders and receivers must remain time synchronized, demanding very accurate clock distribution, for example support for IEEE 1588 transparent clocks in every intermediate node.

In cellular networks from the LTE radio era onward, phase synchronization is needed in addition to frequency synchronization ([TS36300], [TS23401]).

#### 6.1.4. Transport Loss Constraints

Fronthaul and Midhaul networks assume almost error-free transport. Errors can result in a reset of the radio interfaces, which can cause reduced throughput or broken radio connectivity for mobile customers.

For packetized Fronthaul and Midhaul connections packet loss may be caused by BER, congestion, or network failure scenarios. Current tools for elminating packet loss for Fronthaul and Midhaul networks have serious challenges, for example retransmitting lost packets and/or using forward error correction (FEC) to circumvent bit errors is practically impossible due to the additional delay incurred. Using redundant streams for better guarantees for delivery is also practically impossible in many cases due to high bandwidth requirements of Fronthaul and Midhaul networks. Protection switching is also a candidate but current technologies for the path switch are too slow to avoid reset of mobile interfaces.

Fronthaul links are assumed to be symmetric, and all Fronthaul streams (i.e. those carrying radio data) have equal priority and cannot delay or pre-empt each other. This implies that the network must guarantee that each time-sensitive flow meets their schedule.

#### 6.1.5. Security Considerations

Establishing time-sensitive streams in the network entails reserving networking resources for long periods of time. It is important that these reservation requests be authenticated to prevent malicious reservation attempts from hostile nodes (or accidental misconfiguration). This is particularly important in the case where the reservation requests span administrative domains. Furthermore, the reservation information itself should be digitally signed to reduce the risk of a legitimate node pushing a stale or hostile configuration into another networking node.

Note: This is considered important for the security policy of the network, but does not affect the core DetNet architecture and design.

### 6.2. Cellular Radio Networks Today

#### 6.2.1. Fronthaul

Today's Fronthaul networks typically consist of:

- Dedicated point-to-point fiber connection is common
- Proprietary protocols and framings
- Custom equipment and no real networking

Current solutions for Fronthaul are direct optical cables or Wavelength-Division Multiplexing (WDM) connections.

#### 6.2.2. Midhaul and Backhaul

Today's Midhaul and Backhaul networks typically consist of:

- Mostly normal IP networks, MPLS-TP, etc.
- Clock distribution and sync using 1588 and SyncE

Telecommunication networks in the Mid- and Backhaul are already heading towards transport networks where precise time synchronization support is one of the basic building blocks. While the transport networks themselves have practically transitioned to all-IP packet-based networks to meet the bandwidth and cost requirements, highly accurate clock distribution has become a challenge.

In the past, Mid- and Backhaul connections were typically based on Time Division Multiplexing (TDM-based) and provided frequency synchronization capabilities as a part of the transport media. Alternatively other technologies such as Global Positioning System (GPS) or Synchronous Ethernet (SyncE) are used [SyncE].

Both Ethernet and IP/MPLS [RFC3031] (and PseudoWires (PWE) [RFC3985] for legacy transport support) have become popular tools to build and manage new all-IP Radio Access Networks (RANs) [I-D.kh-spring-ip-ran-use-case]. Although various timing and synchronization optimizations have already been proposed and implemented including 1588 PTP enhancements [I-D.ietf-tictoc-1588overmpls] and [I-D.ietf-mpls-residence-time], these solution are not necessarily sufficient for the forthcoming RAN architectures nor do they guarantee the more stringent time-synchronization requirements such as [CPRI].

There are also existing solutions for TDM over IP such as [RFC5087] and [RFC4553], as well as TDM over Ethernet transports such as [RFC5086].

### 6.3. Cellular Radio Networks Future

Future Cellular Radio Networks will be based on a mix of different xHaul networks (xHaul = front-, mid- and backhaul), and future transport networks should be able to support all of them simultaneously. It is already envisioned today that:

- Not all "cellular radio network" traffic will be IP, for example some will remain at Layer 2 (e.g. Ethernet based). DetNet solutions must address all traffic types (Layer 2, Layer 3) with the same tools and allow their transport simultaneously.

- All form of xHaul networks will need some form of DetNet solutions. For example with the advent of 5G some Backhaul traffic will also have DetNet requirements (e.g. traffic belonging to time-critical 5G applications).

We would like to see the following in future Cellular Radio networks:

- Unified standards-based transport protocols and standard networking equipment that can make use of underlying deterministic link-layer services

- Unified and standards-based network management systems and protocols in all parts of the network (including Fronthaul)

New radio access network deployment models and architectures may require time- sensitive networking services with strict requirements on other parts of the network that previously were not considered to be packetized at all. Time and synchronization support are already topical for Backhaul and Midhaul packet networks [MEF] and are becoming a real issue for Fronthaul networks also. Specifically in Fronthaul networks the timing and synchronization requirements can be extreme for packet based technologies, for example, on the order of sub +-20 ns packet delay variation (PDV) and frequency accuracy of +0.002 PPM [Fronthaul].

The actual transport protocols and/or solutions to establish required transport "circuits" (pinned-down paths) for Fronthaul traffic are still undefined. Those are likely to include (but are not limited to) solutions directly over Ethernet, over IP, and using MPLS/ PseudoWire transport.

Even the current time-sensitive networking features may not be sufficient for Fronthaul traffic. Therefore, having specific profiles that take the requirements of Fronthaul into account is desirable [IEEE8021CM].

Interesting and important work for time-sensitive networking has been done for Ethernet [TSNTG], which specifies the use of IEEE 1588 time precision protocol (PTP) [IEEE1588] in the context of IEEE 802.1D and IEEE 802.1Q. [IEEE8021AS] specifies a Layer 2 time synchronizing service, and other specifications such as IEEE 1722 [IEEE1722] specify Ethernet-based Layer-2 transport for time-sensitive streams.

New promising work seeks to enable the transport of time-sensitive fronthaul streams in Ethernet bridged networks [IEEE8021CM]. Analogous to IEEE 1722 there is an ongoing standardization effort to define the Layer-2 transport encapsulation format for transporting radio over Ethernet (RoE) in the IEEE 1904.3 Task Force [IEEE19043].

All-IP RANs and xHhaul networks would benefit from time synchronization and time-sensitive transport services. Although Ethernet appears to be the unifying technology for the transport, there is still a disconnect providing Layer 3 services. The protocol stack typically has a number of layers below the Ethernet Layer 2 that shows up to the Layer 3 IP transport. It is not uncommon that on top of the lowest layer (optical) transport there is the first layer of Ethernet followed one or more layers of MPLS, PseudoWires and/or other tunneling protocols finally carrying the Ethernet layer visible to the user plane IP traffic.

While there are existing technologies to establish circuits through the routed and switched networks (especially in MPLS/PWE space), there is still no way to signal the time synchronization and time-sensitive stream requirements/reservations for Layer-3 flows in a way that addresses the entire transport stack, including the Ethernet layers that need to be configured.

Furthermore, not all "user plane" traffic will be IP. Therefore, the same solution also must address the use cases where the user plane traffic is a different layer, for example Ethernet frames.

There is existing work describing the problem statement [I-D.finn-detnet-problem-statement] and the architecture [I-D.finn-detnet-architecture] for deterministic networking (DetNet) that targets solutions for time-sensitive (IP/transport) streams with deterministic properties over Ethernet-based switched networks.

### 6.4. Cellular Radio Networks Asks

A standard for data plane transport specification which is:

- Unified among all xHauls (meaning that different flows with diverse DetNet requirements can coexist in the same network and traverse the same nodes without interfering with each other)

- Deployed in a highly deterministic network environment

A standard for data flow information models that are:

- Aware of the time sensitivity and constraints of the target networking environment

- Aware of underlying deterministic networking services (e.g., on the Ethernet layer)

## 7. Industrial M2M

### 7.1. Use Case Description

Industrial Automation in general refers to automation of manufacturing, quality control and material processing. In this "machine to machine" (M2M) use case we consider machine units in a plant floor which periodically exchange data with upstream or downstream machine modules and/or a supervisory controller within a local area network.

The actors of M2M communication are Programmable Logic Controllers (PLCs). Communication between PLCs and between PLCs and the supervisory PLC (S-PLC) is achieved via critical control/data streams Figure 8.

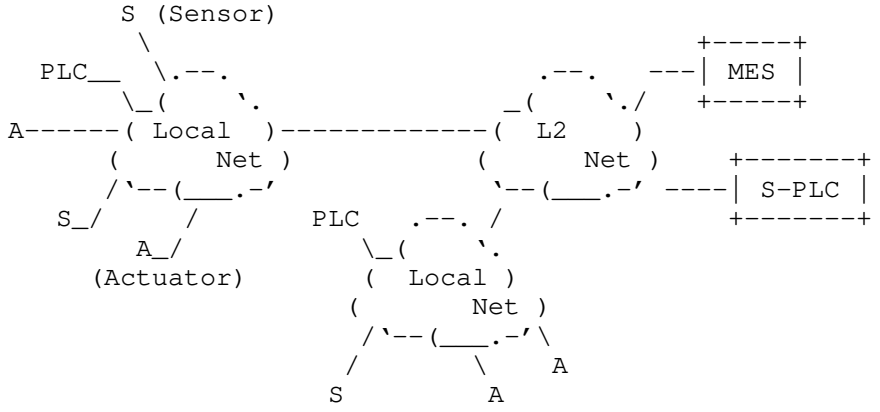



Figure 8: Current Generic Industrial M2M Network Architecture

This use case focuses on PLC-related communications; communication to Manufacturing-Execution-Systems (MESs) are not addressed.

This use case covers only critical control/data streams; non-critical traffic between industrial automation applications (such as communication of state, configuration, set-up, and database communication) are adequately served by currently available prioritizing techniques. Such traffic can use up to 80% of the total bandwidth required. There is also a subset of non-time-critical traffic that must be reliable even though it is not time sensitive.

In this use case the primary need for deterministic networking is to provide end-to-end delivery of M2M messages within specific timing constraints, for example in closed loop automation control. Today this level of determinism is provided by proprietary networking technologies. In addition, standard networking technologies are used to connect the local network to remote industrial automation sites, e.g. over an enterprise or metro network which also carries other types of traffic. Therefore, flows that should be forwarded with deterministic guarantees need to be sustained regardless of the amount of other flows in those networks.

## 7.2. Industrial M2M Communication Today

Today, proprietary networks fulfill the needed timing and availability for M2M networks.

The network topologies used today by industrial automation are similar to those used by telecom networks: Daisy Chain, Ring, Hub and Spoke, and Comb (a subset of Daisy Chain).

PLC-related control/data streams are transmitted periodically and carry either a pre-configured payload or a payload configured during runtime.

Some industrial applications require time synchronization at the end nodes. For such time-coordinated PLCs, accuracy of 1 microsecond is required. Even in the case of "non-time-coordinated" PLCs time sync may be needed e.g. for timestamping of sensor data.

Industrial network scenarios require advanced security solutions. Many of the current industrial production networks are physically separated. Preventing critical flows from be leaked outside a domain is handled today by filtering policies that are typically enforced in firewalls.

#### 7.2.1. Transport Parameters

The Cycle Time defines the frequency of message(s) between industrial actors. The Cycle Time is application dependent, in the range of 1ms - 100ms for critical control/data streams.

Because industrial applications assume deterministic transport for critical Control-Data-Stream parameters (instead of defining latency and delay variation parameters) it is sufficient to fulfill the upper bound of latency (maximum latency). The underlying networking infrastructure must ensure a maximum end-to-end delivery time of messages in the range of 100 microseconds to 50 milliseconds depending on the control loop application.

The bandwidth requirements of control/data streams are usually calculated directly from the bytes-per-cycle parameter of the control loop. For PLC-to-PLC communication one can expect 2 - 32 streams with packet size in the range of 100 - 700 bytes. For S-PLC to PLCs the number of streams is higher - up to 256 streams. Usually no more than 20% of available bandwidth is used for critical control/data streams. In today's networks 1Gbps links are commonly used.

Most PLC control loops are rather tolerant of packet loss, however critical control/data streams accept no more than 1 packet loss per consecutive communication cycle (i.e. if a packet gets lost in cycle "n", then the next cycle ("n+1") must be lossless). After two or more consecutive packet losses the network may be considered to be "down" by the Application.

As network downtime may impact the whole production system the required network availability is rather high (99,999%).

Based on the above parameters we expect that some form of redundancy will be required for M2M communications, however any individual solution depends on several parameters including cycle time, delivery time, etc.

#### 7.2.2. Stream Creation and Destruction

In an industrial environment, critical control/data streams are created rather infrequently, on the order of ~10 times per day / week / month. Most of these critical control/data streams get created at machine startup, however flexibility is also needed during runtime, for example when adding or removing a machine. Going forward as production systems become more flexible, we expect a significant increase in the rate at which streams are created, changed and destroyed.

### 7.3. Industrial M2M Future

We would like to see a converged IP-standards-based network with deterministic properties that can satisfy the timing, security and reliability constraints described above. Today's proprietary networks could then be interfaced to such a network via gateways or, in the case of new installations, devices could be connected directly to the converged network.

For this use case we expect time synchronization accuracy on the order of 1us.

### 7.4. Industrial M2M Asks

- Converged IP-based network
- Deterministic behavior (bounded latency and jitter )
- High availability (presumably through redundancy) (99.999 %)
- Low message delivery time (100us - 50ms)
- Low packet loss (burstless, 0.1-1 %)
- Security (e.g. prevent critical flows from being leaked between physically separated networks)

## 8. Use Case Common Elements

Looking at the use cases collectively, the following common desires for the DetNet-based networks of the future emerge:

- Open standards-based network (replace various proprietary networks, reduce cost, create multi-vendor market)

- Centrally administered (though such administration may be distributed for scale and resiliency)

- Integrates L2 (bridged) and L3 (routed) environments (independent of the Link layer, e.g. can be used with Ethernet, 6TiSCH, etc.)

- Carries both deterministic and best-effort traffic (guaranteed end-to-end delivery of deterministic flows, deterministic flows isolated from each other and from best-effort traffic congestion, unused deterministic BW available to best-effort traffic)

- Ability to add or remove systems from the network with minimal, bounded service interruption (applications include replacement of failed devices as well as plug and play)

- Uses standardized data flow information models capable of expressing deterministic properties (models express device capabilities, flow properties. Protocols for pushing models from controller to devices, devices to controller)

- Scalable size (long distances (many km) and short distances (within a single machine), many hops (radio repeaters, microwave links, fiber links...) and short hops (single machine))

- Scalable timing parameters and accuracy (bounded latency, guaranteed worst case maximum, minimum. Low latency, e.g. control loops may be less than 1ms, but larger for wide area networks)

- High availability (99.9999 percent up time requested, but may be up to twelve 9s)

- Reliability, redundancy (lives at stake)

- Security (from failures, attackers, misbehaving devices - sensitive to both packet content and arrival time)

## 9. Use Cases Explicitly Out of Scope for DetNet

This section contains use case text that has been determined to be outside of the scope of the present DetNet work.

### 9.1. DetNet Scope Limitations

The scope of DetNet is deliberately limited to specific use cases that are consistent with the WG charter, subject to the interpretation of the WG. At the time the DetNet Use Cases were solicited and provided by the authors the scope of DetNet was not clearly defined, and as that clarity has emerged, certain of the use cases have been determined to be outside the scope of the present DetNet work. Such text has been moved into this section to clarify that these use cases will not be supported by the DetNet work.

The text in this section was moved here based on the following "exclusion" principles. Or, as an alternative to moving all such text to this section, some draft text has been modified in situ to reflect these same principles.

The following principles have been established to clarify the scope of the present DetNet work.

- The scope of network addressed by DetNet is limited to networks that can be centrally controlled, i.e. an "enterprise" aka "corporate" network. This explicitly excludes "the open Internet".

- Maintaining synchronized time across a DetNet network is crucial to its operation, however DetNet assumes that time is to be maintained using other means, for example (but not limited to) Precision Time Protocol ([IEEE1588]). A use case may state the accuracy and reliability that it expects from the DetNet network as part of a whole system, however it is understood that such timing properties are not guaranteed by DetNet itself. It is currently an open question as to whether DetNet protocols will include a way for an application to communicate such timing expectations to the network, and if so whether they would be expected to materially affect the performance they would receive from the network as a result.

### 9.2. Internet-based Applications

#### 9.2.1. Use Case Description

There are many applications that communicate across the open Internet that could benefit from guaranteed delivery and bounded latency. The following are some representative examples.

#### 9.2.1.1. Media Content Delivery

Media content delivery continues to be an important use of the Internet, yet users often experience poor quality audio and video due to the delay and jitter inherent in today's Internet.

#### 9.2.1.2. Online Gaming

Online gaming is a significant part of the gaming market, however latency can degrade the end user experience. For example "First Person Shooter" (FPS) games are highly delay-sensitive.

#### 9.2.1.3. Virtual Reality

Virtual reality (VR) has many commercial applications including real estate presentations, remote medical procedures, and so on. Low latency is critical to interacting with the virtual world because perceptual delays can cause motion sickness.

### 9.2.2. Internet-Based Applications Today

Internet service today is by definition "best effort", with no guarantees on delivery or bandwidth.

### 9.2.3. Internet-Based Applications Future

We imagine an Internet from which we will be able to play a video without glitches and play games without lag.

For online gaming, the maximum round-trip delay can be 100ms and stricter for FPS gaming which can be 10-50ms. Transport delay is the dominate part with a 5-20ms budget.

For VR, 1-10ms maximum delay is needed and total network budget is 1-5ms if doing remote VR.

Flow identification can be used for gaming and VR, i.e. it can recognize a critical flow and provide appropriate latency bounds.

### 9.2.4. Internet-Based Applications Asks

- Unified control and management protocols to handle time-critical data flow
- Application-aware flow filtering mechanism to recognize the timing critical flow without doing 5-tuple matching

- o Unified control plane to provide low latency service on Layer-3 without changing the data plane

- o OAM system and protocols which can help to provide E2E-delay sensitive service provisioning

### 9.3. Pro Audio and Video - Digital Rights Management (DRM)

This section was moved here because this is considered a Link layer topic, not direct responsibility of DetNet.

Digital Rights Management (DRM) is very important to the audio and video industries. Any time protected content is introduced into a network there are DRM concerns that must be maintained (see [CONTENT_PROTECTION]). Many aspects of DRM are outside the scope of network technology, however there are cases when a secure link supporting authentication and encryption is required by content owners to carry their audio or video content when it is outside their own secure environment (for example see [DCI]).

As an example, two techniques are Digital Transmission Content Protection (DTCP) and High-Bandwidth Digital Content Protection (HDCP). HDCP content is not approved for retransmission within any other type of DRM, while DTCP may be retransmitted under HDCP. Therefore if the source of a stream is outside of the network and it uses HDCP protection it is only allowed to be placed on the network with that same HDCP protection.

### 9.4. Pro Audio and Video - Link Aggregation

Note: The term "Link Aggregation" is used here as defined by the text in the following paragraph, i.e. not following a more common Network Industry definition. Current WG consensus is that this item won't be directly supported by the DetNet architecture, for example because it implies guarantee of in-order delivery of packets which conflicts with the core goal of achieving the lowest possible latency.

For transmitting streams that require more bandwidth than a single link in the target network can support, link aggregation is a technique for combining (aggregating) the bandwidth available on multiple physical links to create a single logical link of the required bandwidth. However, if aggregation is to be used, the network controller (or equivalent) must be able to determine the maximum latency of any path through the aggregate link.

# 10. Acknowledgments

## 10.1. Pro Audio

This section was derived from draft-gunther-detnet-proaudio-req-01.

The editors would like to acknowledge the help of the following individuals and the companies they represent:

Jeff Koftinoff, Meyer Sound

Jouni Korhonen, Associate Technical Director, Broadcom

Pascal Thubert, CTAO, Cisco

Kieran Tyrrell, Sienda New Media Technologies GmbH

## 10.2. Utility Telecom

This section was derived from draft-wetterwald-detnet-utilities-reqs-02.

Faramarz Maghsoodlou, Ph. D. IoT Connected Industries and Energy Practice Cisco

Pascal Thubert, CTAO Cisco

## 10.3. Building Automation Systems

This section was derived from draft-bas-usecase-detnet-00.

## 10.4. Wireless for Industrial

This section was derived from draft-thubert-6tisch-4detnet-01.

This specification derives from the 6TiSCH architecture, which is the result of multiple interactions, in particular during the 6TiSCH (bi)Weekly Interim call, relayed through the 6TiSCH mailing list at the IETF.

The authors wish to thank: Kris Pister, Thomas Watteyne, Xavier Vilajosana, Qin Wang, Tom Phinney, Robert Assimiti, Michael Richardson, Zhuo Chen, Malisa Vucinic, Alfredo Grieco, Martin Turon, Dominique Barthel, Elvis Vogli, Guillaume Gaillard, Herman Storey, Maria Rita Palattella, Nicola Accettura, Patrick Wetterwald, Pouria Zand, Raghuram Sudhaakar, and Shitanshu Shah for their participation and various contributions.

### 10.5. Cellular Radio

This section was derived from draft-korhonen-detnet-telreq-00.

### 10.6. Industrial M2M

The authors would like to thank Feng Chen and Marcel Kiessling for their comments and suggestions.

### 10.7. Internet Applications and CoMP

This section was derived from draft-zha-detnet-use-case-00.

This document has benefited from reviews, suggestions, comments and proposed text provided by the following members, listed in alphabetical order: Jing Huang, Junru Lin, Lehong Niu and Oilver Huang.

## 11. Informative References

[ACE] IETF, "Authentication and Authorization for Constrained Environments", <https://datatracker.ietf.org/doc/charter-ietf-ace/>.

[bacnetip] ASHRAE, "Annex J to ANSI/ASHRAE 135-1995 - BACnet/IP", January 1999.

[CCAMP] IETF, "Common Control and Measurement Plane", <https://datatracker.ietf.org/doc/charter-ietf-ccamp/>.

[CoMP] NGMN Alliance, "RAN EVOLUTION PROJECT COMP EVALUATION AND ENHANCEMENT", NGMN Alliance NGMN_RANEV_D3_CoMP_Evaluation_and_Enhancement_v2.0, March 2015, <https://www.ngmn.org/uploads/media/NGMN_RANEV_D3_CoMP_Evaluation_and_Enhancement_v2.0.pdf>.

[CONTENT_PROTECTION] Olsen, D., "1722a Content Protection", 2012, <http://grouper.ieee.org/groups/1722/contributions/2012/avtp_dolsen_1722a_content_protection.pdf>.

[CPRI] CPRI Cooperation, "Common Public Radio Interface (CPRI); Interface Specification", CPRI Specification V6.1, July 2014, <http://www.cpri.info/downloads/CPRI_v_6_1_2014-07-01.pdf>.

[CPRI-transp]
CPRI TWG, "CPRI requirements for Ethernet Fronthaul", November 2015, <http://www.ieee802.org/1/files/public/docs2015/cm-CPRI-requirements-1115-v01.pdf>.

[DCI] Digital Cinema Initiatives, LLC, "DCI Specification, Version 1.2", 2012, <http://www.dcimovies.com/>.

[DICE] IETF, "DTLS In Constrained Environments", <https://datatracker.ietf.org/doc/charter-ietf-dice/>.

[EA12] Evans, P. and M. Annunziata, "Industrial Internet: Pushing the Boundaries of Minds and Machines", November 2012.

[ESPN_DC2]
Daley, D., "ESPN's DC2 Scales AVB Large", 2014, <http://sportsvideo.org/main/blog/2014/06/espns-dc2-scales-avb-large>.

[flnet] Japan Electrical Manufacturers' Association, "JEMA 1479 - English Edition", September 2012.

[Fronthaul]
Chen, D. and T. Mustala, "Ethernet Fronthaul Considerations", IEEE 1904.3, February 2015, <http://www.ieee1904.org/3/meeting_archive/2015/02/tf3_1502_che n_1a.pdf>.

[HART] www.hartcomm.org, "Highway Addressable remote Transducer, a group of specifications for industrial process and control devices administered by the HART Foundation".

[I-D.finn-detnet-architecture]
Finn, N., Thubert, P., and M. Teener, "Deterministic Networking Architecture", draft-finn-detnet-architecture-04 (work in progress), March 2016.

[I-D.finn-detnet-problem-statement]
Finn, N. and P. Thubert, "Deterministic Networking Problem Statement", draft-finn-detnet-problem-statement-05 (work in progress), March 2016.

[I-D.ietf-6tisch-6top-interface]
Wang, Q. and X. Vilajosana, "6TiSCH Operation Sublayer (6top) Interface", draft-ietf-6tisch-6top-interface-04 (work in progress), July 2015.

[I-D.ietf-6tisch-architecture]
Thubert, P., "An Architecture for IPv6 over the TSCH mode of IEEE 802.15.4", draft-ietf-6tisch-architecture-10 (work in progress), June 2016.

[I-D.ietf-6tisch-coap]
Sudhaakar, R. and P. Zand, "6TiSCH Resource Management and Interaction using CoAP", draft-ietf-6tisch-coap-03 (work in progress), March 2015.

[I-D.ietf-6tisch-terminology]
Palattella, M., Thubert, P., Watteyne, T., and Q. Wang, "Terminology in IPv6 over the TSCH mode of IEEE 802.15.4e", draft-ietf-6tisch-terminology-07 (work in progress), March 2016.

[I-D.ietf-ipv6-multilink-subnets]
Thaler, D. and C. Huitema, "Multi-link Subnet Support in IPv6", draft-ietf-ipv6-multilink-subnets-00 (work in progress), July 2002.

[I-D.ietf-mpls-residence-time]
Mirsky, G., Ruffini, S., Gray, E., Drake, J., Bryant, S., and S. Vainshtein, "Residence Time Measurement in MPLS network", draft-ietf-mpls-residence-time-09 (work in progress), April 2016.

[I-D.ietf-roll-rpl-industrial-applicability]
Phinney, T., Thubert, P., and R. Assimiti, "RPL applicability in industrial networks", draft-ietf-roll-rpl-industrial-applicability-02 (work in progress), October 2013.

[I-D.ietf-tictoc-1588overmpls]
Davari, S., Oren, A., Bhatia, M., Roberts, P., and L. Montini, "Transporting Timing messages over MPLS Networks", draft-ietf-tictoc-1588overmpls-07 (work in progress), October 2015.

[I-D.kh-spring-ip-ran-use-case]
Khasnabish, B., hu, f., and L. Contreras, "Segment Routing in IP RAN use case", draft-kh-spring-ip-ran-use-case-02 (work in progress), November 2014.

[I-D.svshah-tsvwg-deterministic-forwarding]
Shah, S. and P. Thubert, "Deterministic Forwarding PHB", draft-svshah-tsvwg-deterministic-forwarding-04 (work in progress), August 2015.

[I-D.thubert-6lowpan-backbone-router]
Thubert, P., "6LoWPAN Backbone Router", draft-thubert-6lowpan-backbone-router-03 (work in progress), February 2013.

[I-D.wang-6tisch-6top-sublayer]
Wang, Q. and X. Vilajosana, "6TiSCH Operation Sublayer (6top)", draft-wang-6tisch-6top-sublayer-04 (work in progress), November 2015.

[IEC61850-90-12]
TC57 WG10, IEC., "IEC 61850-90-12 TR: Communication networks and systems for power utility automation - Part 90-12: Wide area network engineering guidelines", 2015.

[IEC62439-3:2012]
TC65, IEC., "IEC 62439-3: Industrial communication networks - High availability automation networks - Part 3: Parallel Redundancy Protocol (PRP) and High-availability Seamless Redundancy (HSR)", 2012.

[IEEE1588]
IEEE, "IEEE Standard for a Precision Clock Synchronization Protocol for Networked Measurement and Control Systems", IEEE Std 1588-2008, 2008, <http://standards.ieee.org/findstds/standard/1588-2008.html>.

[IEEE1722]
IEEE, "1722-2011 - IEEE Standard for Layer 2 Transport Protocol for Time Sensitive Applications in a Bridged Local Area Network", IEEE Std 1722-2011, 2011, <http://standards.ieee.org/findstds/standard/1722-2011.html>.

[IEEE19043]
IEEE Standards Association, "IEEE 1904.3 TF", IEEE 1904.3, 2015, <http://www.ieee1904.org/3/tf3_home.shtml>.

[IEEE802.1TSNTG]
IEEE Standards Association, "IEEE 802.1 Time-Sensitive Networks Task Group", March 2013, <http://www.ieee802.org/1/pages/avbridges.html>.

[IEEE802154]
IEEE standard for Information Technology, "IEEE std. 802.15.4, Part. 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks".

[IEEE802154e]
IEEE standard for Information Technology, "IEEE standard for Information Technology, IEEE std. 802.15.4, Part. 15.4: Wireless Medium Access Control (MAC) and Physical Layer (PHY) Specifications for Low-Rate Wireless Personal Area Networks, June 2011 as amended by IEEE std. 802.15.4e, Part. 15.4: Low-Rate Wireless Personal Area Networks (LR-WPANs) Amendment 1: MAC sublayer", April 2012.

[IEEE8021AS]
IEEE, "Timing and Synchronizations (IEEE 802.1AS-2011)", IEEE 802.1AS-2001, 2011, <http://standards.ieee.org/getIEEE802/download/802.1AS-2011.pdf>.

[IEEE8021CM]
Farkas, J., "Time-Sensitive Networking for Fronthaul", Unapproved PAR, PAR for a New IEEE Standard; IEEE P802.1CM, April 2015, <http://www.ieee802.org/1/files/public/docs2015/new-P802-1CM-dr aft-PAR-0515-v02.pdf>.

[IEEE8021TSN]
IEEE 802.1, "The charter of the TG is to provide the specifications that will allow time-synchronized low latency streaming services through 802 networks.", 2016, <http://www.ieee802.org/1/pages/tsn.html>.

[IETFDetNet]
IETF, "Charter for IETF DetNet Working Group", 2015, <https://datatracker.ietf.org/wg/detnet/charter/>.

[ISA100]
ISA/ANSI, "ISA100, Wireless Systems for Automation", <https://www.isa.org/isa100/>.

[ISA100.11a]
ISA/ANSI, "Wireless Systems for Industrial Automation: Process Control and Related Applications - ISA100.11a-2011 - IEC 62734", 2011, <http://www.isa.org/Community/SP100WirelessSystemsforAutomation>.

[ISO7240-16]
ISO, "ISO 7240-16:2007 Fire detection and alarm systems -- Part 16: Sound system control and indicating equipment", 2007, <http://www.iso.org/iso/catalogue_detail.htm?csnumber=42978>.

[knx] KNX Association, "ISO/IEC 14543-3 - KNX", November 2006.

[lontalk] ECHELON, "LonTalk(R) Protocol Specification Version 3.0", 1994.

[LTE-Latency]
Johnston, S., "LTE Latency: How does it compare to other technologies", March 2014, <http://opensignal.com/blog/2014/03/10/lte-latency-how-does-it-compare-to-other-technologies>.

[MEF] MEF, "Mobile Backhaul Phase 2 Amendment 1 -- Small Cells", MEF 22.1.1, July 2014, <http://www.mef.net/Assets/Technical_Specifications/PDF/MEF_22.1.1.pdf>.

[METIS] METIS, "Scenarios, requirements and KPIs for 5G mobile and wireless system", ICT-317669-METIS/D1.1 ICT-317669-METIS/D1.1, April 2013, <https://www.metis2020.com/wp-content/uploads/deliverables/METIS_D1.1_v1.pdf>.

[modbus] Modbus Organization, "MODBUS APPLICATION PROTOCOL SPECIFICATION V1.1b", December 2006.

[net5G] Ericsson, "5G Radio Access, Challenges for 2020 and Beyond", Ericsson white paper wp-5g, June 2013, <http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf>.

[NGMN] NGMN Alliance, "5G White Paper", NGMN 5G White Paper v1.0, February 2015, <https://www.ngmn.org/uploads/media/NGMN_5G_White_Paper_V1_0.pdf>.

[NGMN-fronth]
NGMN Alliance, "Fronthaul Requirements for C-RAN", March 2015, <https://www.ngmn.org/uploads/media/NGMN_RANEV_D1_C-RAN_Fronthaul_Requirements_v1.0.pdf>.

[PCE] IETF, "Path Computation Element", <https://datatracker.ietf.org/doc/charter-ietf-pce/>.

[profibus]
IEC, "IEC 61158 Type 3 - Profibus DP", January 2001.

[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>.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, DOI 10.17487/RFC2460, December 1998, <http://www.rfc-editor.org/info/rfc2460>.

[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, <http://www.rfc-editor.org/info/rfc2474>.

[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>.

[RFC3393] Demichelis, C. and P. Chimento, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM)", RFC 3393, DOI 10.17487/RFC3393, November 2002, <http://www.rfc-editor.org/info/rfc3393>.

[RFC3444] Pras, A. and J. Schoenwaelder, "On the Difference between Information Models and Data Models", RFC 3444, DOI 10.17487/RFC3444, January 2003, <http://www.rfc-editor.org/info/rfc3444>.

[RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, DOI 10.17487/RFC3972, March 2005, <http://www.rfc-editor.org/info/rfc3972>.

[RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation Edge-to-Edge (PWE3) Architecture", RFC 3985, DOI 10.17487/RFC3985, March 2005, <http://www.rfc-editor.org/info/rfc3985>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <http://www.rfc-editor.org/info/rfc4291>.

[RFC4553] Vainshtein, A., Ed. and YJ. Stein, Ed., "Structure-Agnostic Time Division Multiplexing (TDM) over Packet (SAToP)", RFC 4553, DOI 10.17487/RFC4553, June 2006, <http://www.rfc-editor.org/info/rfc4553>.

[RFC4903] Thaler, D., "Multi-Link Subnet Issues", RFC 4903, DOI 10.17487/RFC4903, June 2007, <http://www.rfc-editor.org/info/rfc4903>.

[RFC4919] Kushalnagar, N., Montenegro, G., and C. Schumacher, "IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs): Overview, Assumptions, Problem Statement, and Goals", RFC 4919, DOI 10.17487/RFC4919, August 2007, <http://www.rfc-editor.org/info/rfc4919>.

[RFC5086] Vainshtein, A., Ed., Sasson, I., Metz, E., Frost, T., and P. Pate, "Structure-Aware Time Division Multiplexed (TDM) Circuit Emulation Service over Packet Switched Network (CESoPSN)", RFC 5086, DOI 10.17487/RFC5086, December 2007, <http://www.rfc-editor.org/info/rfc5086>.

[RFC5087] Stein, Y(J)., Shashoua, R., Insler, R., and M. Anavi, "Time Division Multiplexing over IP (TDMoIP)", RFC 5087, DOI 10.17487/RFC5087, December 2007, <http://www.rfc-editor.org/info/rfc5087>.

[RFC6282] Hui, J., Ed. and P. Thubert, "Compression Format for IPv6 Datagrams over IEEE 802.15.4-Based Networks", RFC 6282, DOI 10.17487/RFC6282, September 2011, <http://www.rfc-editor.org/info/rfc6282>.

[RFC6550] Winter, T., Ed., Thubert, P., Ed., Brandt, A., Hui, J., Kelsey, R., Levis, P., Pister, K., Struik, R., Vasseur, JP., and R. Alexander, "RPL: IPv6 Routing Protocol for Low-Power and Lossy Networks", RFC 6550, DOI 10.17487/RFC6550, March 2012, <http://www.rfc-editor.org/info/rfc6550>.

[RFC6551] Vasseur, JP., Ed., Kim, M., Ed., Pister, K., Dejean, N., and D. Barthel, "Routing Metrics Used for Path Calculation in Low-Power and Lossy Networks", RFC 6551, DOI 10.17487/RFC6551, March 2012, <http://www.rfc-editor.org/info/rfc6551>.

[RFC6775] Shelby, Z., Ed., Chakrabarti, S., Nordmark, E., and C. Bormann, "Neighbor Discovery Optimization for IPv6 over Low-Power Wireless Personal Area Networks (6LoWPANs)", RFC 6775, DOI 10.17487/RFC6775, November 2012, <http://www.rfc-editor.org/info/rfc6775>.

[RFC7554] Watteyne, T., Ed., Palattella, M., and L. Grieco, "Using IEEE 802.15.4e Time-Slotted Channel Hopping (TSCH) in the Internet of Things (IoT): Problem Statement", RFC 7554, DOI 10.17487/RFC7554, May 2015, <http://www.rfc-editor.org/info/rfc7554>.

[SRP_LATENCY] Gunther, C., "Specifying SRP Latency", 2014, <http://www.ieee802.org/1/files/public/docs2014/cc-cgunther-acceptable-latency-0314-v01.pdf>.

[STUDIO_IP] Mace, G., "IP Networked Studio Infrastructure for Synchronized & Real-Time Multimedia Transmissions", 2007, <http://www.ieee802.org/1/files/public/docs2047/avb-mace-ip-networked-studio-infrastructure-0107.pdf>.

[SyncE] ITU-T, "G.8261 : Timing and synchronization aspects in packet networks", Recommendation G.8261, August 2013, <http://www.itu.int/rec/T-REC-G.8261>.

[TEAS] IETF, "Traffic Engineering Architecture and Signaling", <https://datatracker.ietf.org/doc/charter-ietf-teas/>.

[TS23401] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", 3GPP TS 23.401 10.10.0, March 2013.

[TS25104] 3GPP, "Base Station (BS) radio transmission and reception (FDD)", 3GPP TS 25.104 3.14.0, March 2007.

[TS36104] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Base Station (BS) radio transmission and reception", 3GPP TS 36.104 10.11.0, July 2013.

[TS36133] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Requirements for support of radio resource management", 3GPP TS 36.133 12.7.0, April 2015.

[TS36211] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation", 3GPP TS 36.211 10.7.0, March 2013.

[TS36300] 3GPP, "Evolved Universal Terrestrial Radio Access (E-UTRA) and Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Overall description; Stage 2", 3GPP TS 36.300 10.11.0, September 2013.

[TSNTG] IEEE Standards Association, "IEEE 802.1 Time-Sensitive Networks Task Group", 2013, <http://www.IEEE802.org/1/pages/avbridges.html>.

[UHD-video] Holub, P., "Ultra-High Definition Videos and Their Applications over the Network", The 7th International Symposium on VICTORIES Project PetrHolub_presentation, October 2014, <http://www.aist-victories.org/jp/7th_sympo_ws/PetrHolub_presentation.pdf>.

[WirelessHART] www.hartcomm.org, "Industrial Communication Networks - Wireless Communication Network and Communication Profiles - WirelessHART - IEC 62591", 2010.

## Authors' Addresses

Ethan Grossman (editor)
Dolby Laboratories, Inc.
1275 Market Street
San Francisco, CA 94103
USA

Phone: +1 415 645 4726
Email: ethan.grossman@dolby.com
URI: http://www.dolby.com

Craig Gunther
Harman International
10653 South River Front Parkway
South Jordan, UT 84095
USA

Phone: +1 801 568-7675
Email: craig.gunther@harman.com
URI: http://www.harman.com

Pascal Thubert
Cisco Systems, Inc
Building D
45 Allee des Ormes - BP1200
MOUGINS - Sophia Antipolis 06254
FRANCE

Phone: +33 497 23 26 34
Email: pthubert@cisco.com

Patrick Wetterwald
Cisco Systems
45 Allees des Ormes
Mougins 06250
FRANCE

Phone: +33 4 97 23 26 36
Email: pwetterw@cisco.com

Jean Raymond
Hydro-Quebec
1500 University
Montreal H3A3S7
Canada

Phone: +1 514 840 3000
Email: raymond.jean@hydro.qc.ca

Jouni Korhonen
Broadcom Corporation
3151 Zanker Road
San Jose, CA 95134
USA

Email: jouni.nospam@gmail.com

Yu Kaneko
Toshiba
1 Komukai-Toshiba-cho, Saiwai-ku, Kasasaki-shi
Kanagawa, Japan

Email: yu1.kaneko@toshiba.co.jp

Subir Das
Applied Communication Sciences
150 Mount Airy Road, Basking Ridge
New Jersey, 07920, USA

Email: sdas@appcomsci.com

Yiyong Zha
Huawei Technologies

Email: zhayiyong@huawei.com

Balazs Varga
Ericsson
Konyves Kalman krt. 11/B
Budapest 1097
Hungary

Email: balazs.a.varga@ericsson.com

Janos Farkas
Ericsson
Konyves Kalman krt. 11/B
Budapest 1097
Hungary

Email: janos.farkas@ericsson.com

Franz-Josef Goetz
Siemens
Gleiwitzerstr. 555
Nurnberg 90475
Germany

Email: franz-josef.goetz@siemens.com

Juergen Schmitt
Siemens
Gleiwitzerstr. 555
Nurnberg 90475
Germany

Email: juergen.jues.schmitt@siemens.com

DetNet
Internet-Draft
Intended status: Standards Track
Expires: January 9, 2017

B. Varga, Ed.
J. Farkas
Ericsson

July 08, 2016

# DetNet Service Model
## draft-varga-detnet-service-model-00

Abstract

This document describes the service model for scenarios requiring deterministic / time sensitive networking.

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 January 9, 2017.

Table of Contents



## 1. Introduction

Deterministic Networking service provides a capability to carry specified data flow, whether unicast or multicast, for an application with constrained requirements towards network performance, e.g. low packet loss rate and/or latency. During the discussion of detnet use-cases, detnet architecture and various related networking scenarios several confusions have been arrised due to different service model interpretations. This document defines service reference points, service components and proposes naming for service scenarios to achieve common understanding of the detnet service model.

## 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 [RFC2119].

The lowercase forms with an initial capital "Must", "Must Not", "Shall", "Shall Not", "Should", "Should Not", "May", and "Optional" in this document are to be interpreted in the sense defined in [RFC2119], but are used where the normative behavior is defined in documents published by SDOs other than the IETF.

## 3. Terminology and Definitions

Additional terms to [draft-data-plane] and [draft-arch] used/ described in this draft .

App-flow: Data flow between the applications requiring deterministic transport.

DetLink: Direct link between two entities (node/end-system) used for deterministic transport.

DetNet AC: Attachment Circuit of a DetNet transport service for a DetNet-flow.

DetNet-flow: Data flow requiring deterministic transport between two DetNet-UNIs.

DetNet-UNI: UNI of an Edge/Relay node to provide deterministic service for a connected node/end-system.

DetNetwork: Transport network between DetNet-UNIs.

Native AC: Attachment Circuit of a DetNet transport service for an App-flow.

## 4. End-systems connected to DetNet

Deterministic transport is required by time/loss sensitive application(s) running on an End-system during communication with its peer(s). Such a data exchange has various requirements on delay and/ or loss parameters. The native data flow between the source/sink End-Systems is called as application-flow (app-flow) as shown in Figure 1. The traffic characteristics of an app-flow can be CBR or VBR and can have L1 or L2 or L3 format (e.g., TDM, Ethernet, IP).

[Note: Interworking function for L1 application-flows is out-of-scope in this document and therefore not depicted on figures.]

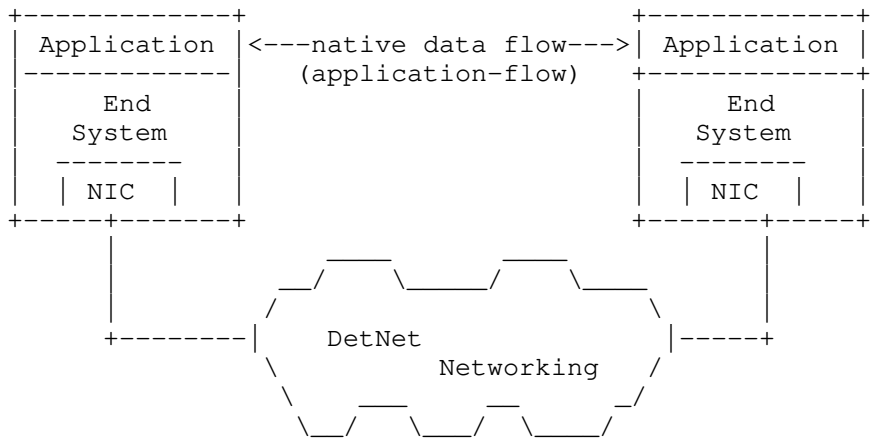



Figure 1: End-systems connected to DetNet

End-system(s) may or may not be directly connected to the DetNet transport network. End-systems may or may not contain DetNet specific functionalities and are referred as "DetNet aware End-sytem" or "DetNet unaware End-system" respectively [draft-data-plane]. (Note: [draft-data-plane] refers to "DetNet unaware end-systems" as "TSN End-system")

- "DetNet aware End-system" has the same forwarding paradigm as the DetNet transport network and it creates the DetNet-flow from the app-flow. DetNet aware End-system is connected via "DetNet AC" to the DetNet transport network.

- "DetNet unaware End-system" originates a native data flow (app-flow) from which an Edge node creates a DetNet-flow (with proper Flow-ID and Seq-num attributes) by encapsulating native data flow according to the forwarding paradigm of the transport network. DetNet unaware End-system is connected via "Native AC" to the DetNet transport network.

These end-systems are shown in Figure 2

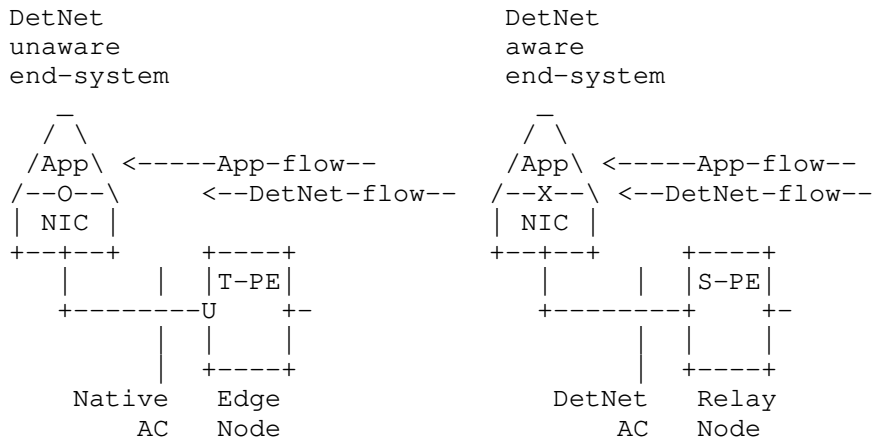



Figure 2: DetNet aware/unaware End-systems

## 5. DetNet service model

### 5.1. Service parameters

The DetNet service can be defined as a service, that provides a capability to carry specified data flow, whether unicast or multicast, for an application with constrained requirements towards network performance, e.g. low packet loss rate and/or latency.

Delay and loss parameters are somewhat correlated as too late arrival of a packet can be treated as lost. However not all applications require hard limits on both parameters (delay and loss). For example, some real-time applications allow graceful degradation if loss happens (e.g., samples based processing, media distribution). Some others may require high bandwidth connections that makes the usage of techniques like flow duplication economically challenging or even impossible. Some applications may not tolerate loss, but are not delay sensitive (e.g., bufferless sensors).

Primary transport service attributes for DetNet transport are:

- Bandwidth parameter(s),
- Delay parameter(s),
- Loss parameter(s).

Time/Loss sensitive applications may have somewhat special requirements especially for loss (e.g. no loss in two consecutives communication cycles; very low outage time, etc.).

## 5.2. Service overview

The figure below shows the DetNet service related reference points and components (Figure 3).

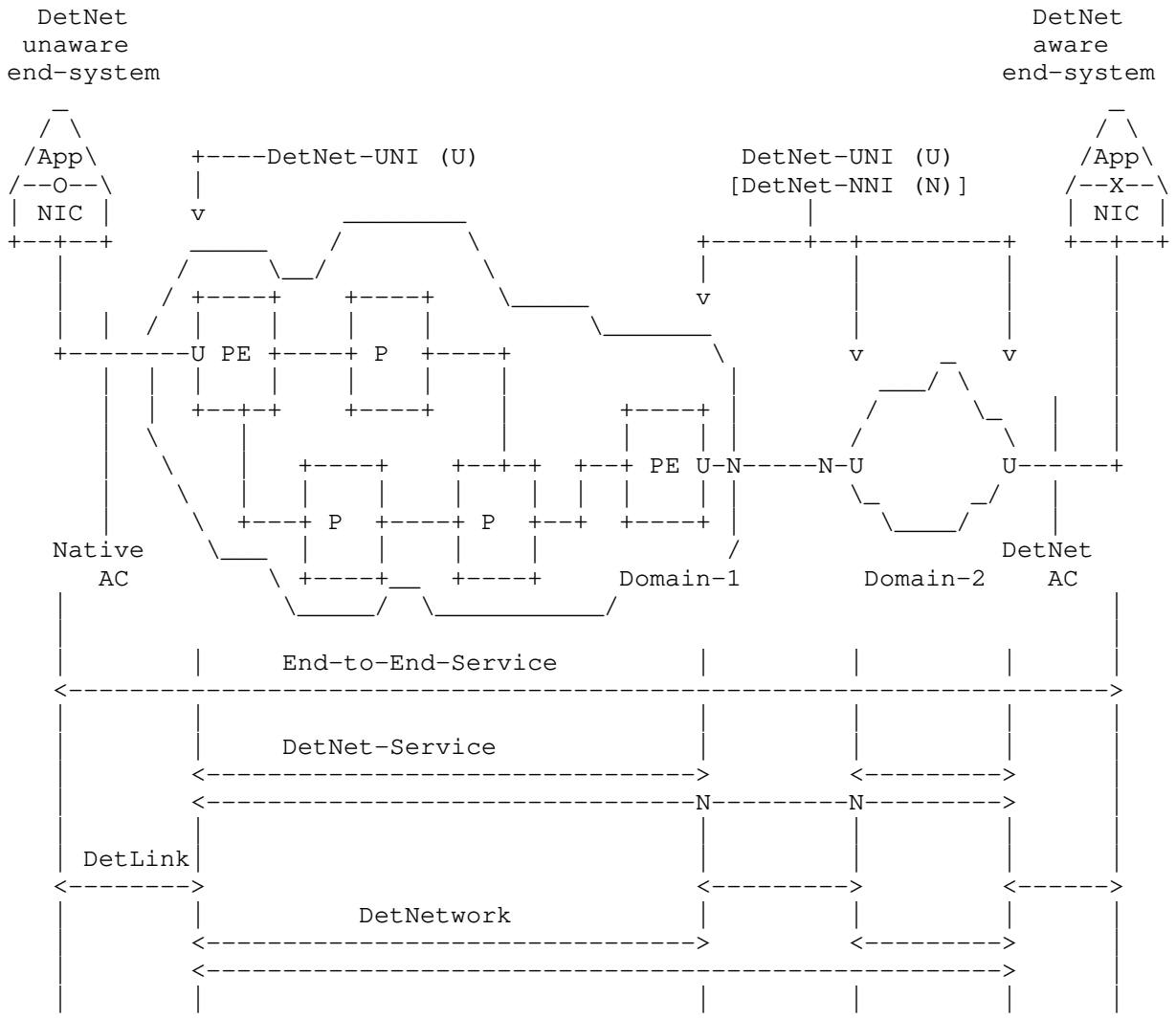



Figure 3: DetNet Service Reference Model

## 5.3. Reference Points

From service model design perspective a fundamental question is the location of the service endpoints, i.e., where the service starts or ends. Two reference points can be distinguished for all DetNet use-cases:

- App-flow endpoints: end-system's internal reference point.

- DetNet-UNI: edge node UNI interface of a domain.

App-flow endpoints (depicted as "O" and "X" on Figure 3) is a more challenging point from control perspective as it is an internal reference point. It is providing access to deterministic transport for the native data flow (app-flow).

A DetNet-UNI (depicted as "U" on Figure 3) is assumed in this document to be a packet based reference point and provides connectivity over the packet network. A DetNet-UNI may add networking technology specific encapsulation to the app-flow and transport it as a DetNet-Flow over the network. There are many similarities regarding the functions of an app-flow endpoint ("X") of an DetNet aware endsystem and DetNet-UNI but there may be some differences. For example in-order delivery is expected over the app-flow endpoints ("O/X"), whereas it is considered as optional over the DetNet-UNI.

## 5.4. Service scenarios

Using the above defined reference points two major service scenarios can be created:

- End-to-End-Service: the service reaches out to final source/sink nodes, so it is an e2e service between application hosting devices (end-systems).

- DetNet-Service: the service connects networking islands, so it is a service between the borders of network domain(s).

End-to-End-Service is defined between app-flow endpoints, whereas DetNet-Service between DetNet-UNIs. That allows peering of same layers/functions.

## 5.5. Data flows

For unambiguous references two detnet data flows are distinguished:

- App-flow: data flow requiring deterministic transport between two app-flow endpoints, data format is application specific (e.g., bit stream, directly mapped in Ethernet frames, etc.).

- DetNet-flow: data flow requiring deterministic transport between two DetNet-UNIs. Data format may be chaged at the DetNet-UNI to allow simple flow recognition/transport/etc. during forwarding between DetNet-UNIs (e.g., on Edge Nodes by adding further encapsulation to the App-flow including new domain specific Flow-ID and Seq-num attributes) .

[Note: In some network scenarios App-flow and DetNet-flow format might be identical e.g., if the end-system provides an encapsulation, that contains all information needed by detnet functionalities (e.g., RTP based App-flow trasported over a native IP network). In other scenarios their encapsulation format might differ significantly e.g., CPRI IQ data mapped directly to Ethernet frames which have to be transported over an MPLS based PSN.]

### 5.6. Service components/segments

As a reference to service components/segments the follwoing building blocks are used:

- DetLink: direct link between two entities (node/end-system) used for deterministic transport.

- DetNetwork: network between DetNet-UNIs

Using DetLink and DetNetwork component/segments any detnet service scenario can be described. For example the service between the App-flow endpoints on Figure 3 can be composed as a DetLink-1 (between end-system on the left and the edge node of domain-1) + DetNetwork-1 (of domain-1) + DetLink-2 (between domain-1 and domain-2) + DetNetwork-2 (of domain-2) + DetLink-3 (between edge node of domain-2 and end-system on the right).

## 6. DetNet service instances

### 6.1. Local attributes used by DetNet functions

The three DetNet functions (Bandwidth reservation and enforcement, Explicit routes, Packet loss protection) require two data flow related attributes to work properly:

- Flow-ID and

- Sequence number (Seq-Num).

These attributes are local to DetNet nodes and are extracted from the ingress packets of the node [draft-arch]. Flow-ID is used by all the three DetNet functions, but sequence number is used only by the duplicate elimination functionality.

Flow-ID must be unique per network domain. Its encoding format is specific to the forwarding paradigm of the domain and to the capabilities of intermediate nodes to identify data flows. For example in case of "PW over MPLS" one option is to construct the Flow-ID by the PW label and the LSP label (denoted as [PW-label;LSP-

label]). In such a case intermediate P nodes have to check all labels to identify a DetNet-flow. An other possible option is to use a dedicated LSP per data flow so the LSP label itself can be used as a Flow-ID (denoted as [LSP-label]). In such a case the intermediate P nodes do not have to check the whole label stack to recognize a data flow (DetNet-flow).

## 6.2. Service instance for DetNet data flows

The DetNet network reference model is shown in Figure 4 for a DetNet-Service scenario (i.e. between two DetNet-UNIs). In this figure the end-systems ("A" and "B") are connected directly to the edge nodes of the PSN ("PE1" and "PE2"). The data flow specific attachment circuits ("AC-A" and "AC-B") are terminated in the flow specific service instance ("SI-1" and "SI-2"). A PSN tunnel is used to provide connectivity between the service instances. The encapsulations used over the PSN tunnel are described in [draft-data-plane].

This PSN tunnel is used to transport exclusivelly the DetNet data flow packets between "SI-1" and "SI-2". The service instances are configured to implement a flow specific routing or bridging function depending on what connectivity (L2 or L3) the participating end-systems require. The service instance and the PSN tunnel may or may not be shared by multiple DetNet data flows. Sharing the service istance by multiple DetNet-flows requires properly populated forwarding tables of the service instance.

Serving regular traffic and DetNet data flows by the same service instance is out-of-scope in this draft, but some related thoughts are described in the annex.

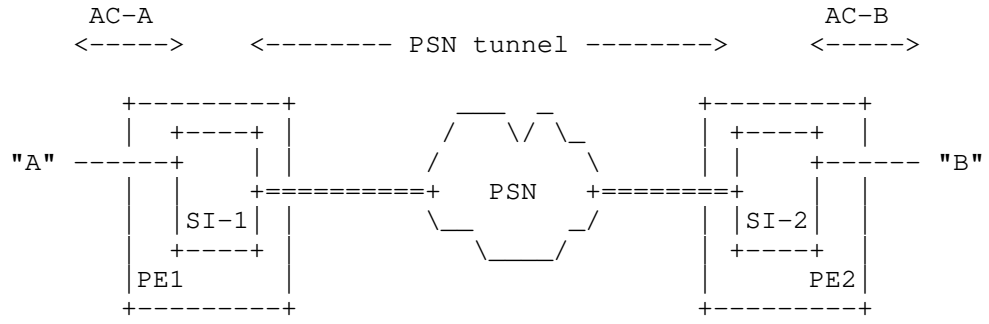



Figure 4: DetNet network reference model

[Note: There are differences in the usage of a "packet PW" for DetNet traffic compared to the network model described in [rfc6658]. In the DetNet scenario the packet PW is used exclusivelly by the DetNet data flows, whereas RFC6658 states: "The packet PW appears as a single point-to-point link to the client layer. Network-layer adjacency formation and maintenance between the client equipments will follow the normal practice needed to support the required relationship in the client layer ... This packet pseudowire is used to transport all of the required layer 2 and layer 3 protocols between LSR1 and LSR2".]

# 7. DetNet data flows over multiple technology domains

## 7.1. Flow attribute mappings between layers

Transport of DetNet data flows over multiple technology domains may require that e.g., lower layers are aware of specific flows at higher layers. Such an "exporting of flow identification" [see draft-arch section 4.7] is needed each time when the forwarding paradigm is changed on the transport path (e.g., two LSRs are interconnected by a L2 bridged domain, etc.). The three main forwarding methods considered for deterministic networking are:

- IP routing
- MPLS label switching
- Ethernet bridging

The simplest solution for generalized flow identification could be to define a unique Flow-ID triplet per DetNet data flow (e.g., [IP: "IPv6-flow-label"+"IPv6-address"; MPLS: "PW-label"+"LSP-label"; Ethernet: "VLAN-ID"+"MAC-address"). This triplet can be used by the DetNet encoding function of technology border nodes (where forwarding paradigm changes) to adapt to capabilities of the next hop node. They push a further (forwarding paradigm specific) Flow-ID to packets ensuring that flows can be easily recognized by domain internal nodes. This additional Flow-ID might be removed when the packet leaves a given technology domain.

[Note: Seq-num attribute may require a similar functionality at technology border nodes.]

The additional (domain specific) Flow-ID can be

- created by a domain specific function or
- derived from the original Flow-ID of the app-flow

so, that it must be unique inside the given domain. Please note, that the original Flow-ID of the app-flow is still present in the packet, but transport nodes may lack the function to recognize it, that's why the additional Flow-ID is added (pushed).

## 7.2. Flow-ID mappings examples

IP nodes and MPLS nodes are assumed to be configured to push such an additional (domain specific) Flow-ID when sending traffic to an Ethernet switch (as shown in the examples below).

Figure 5 below shows a scenario where an IP end-system ("IP-A") is connected via two Ethernet switches ("ETH-n") to an IP router ("IP-1").

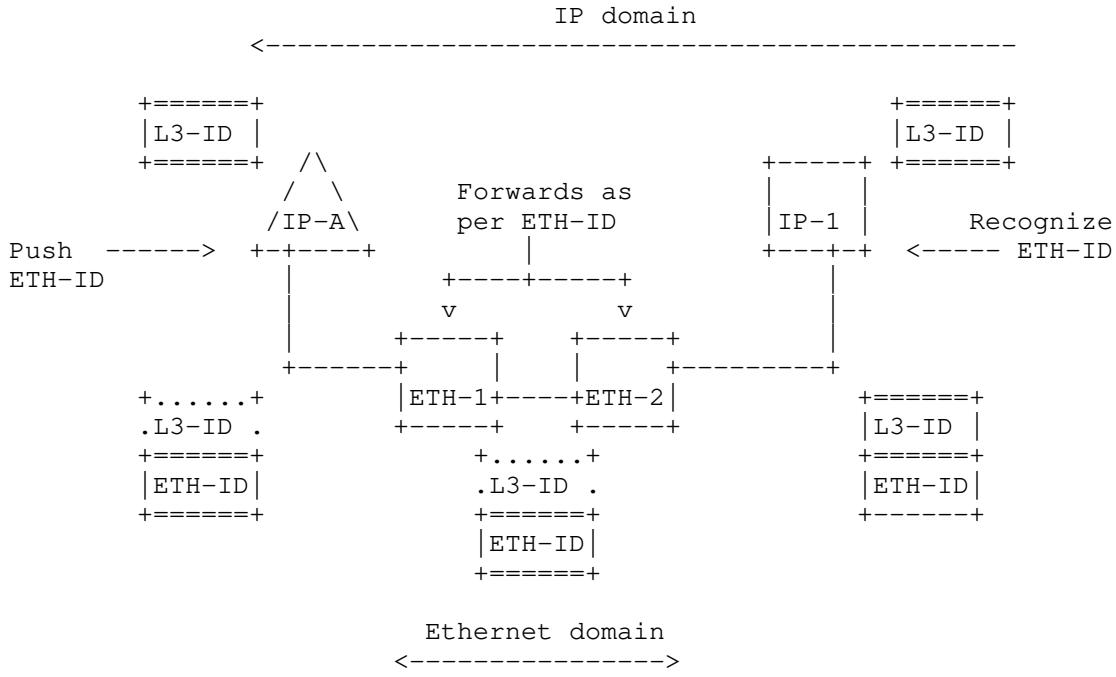



Figure 5: IP nodes interconnected by an Ethernet domain

"IP-A" uses the original App-flow specific ID ("L3-ID"), but as it is connected to an Ethernet domain it has to push also an Ethernet-domain specific flow-ID ("VLAN+multicast-MAC", referred as "ETH-ID") before sending the packet to "ETH-1". The ethernet switch "ETH-1" can recognize the data flow based on the "ETH-ID" and it does forwarding towards "ETH-2". "ETH-2" switches the packet towards the

IP router. "IP-1" must be configured to receive the Ethernet Flow-ID specific multicast stream, but (as it is an L3 node) it decodes the data flow ID based on the "L3-ID" fields of the received packet.

Figure 6 below shows a scenario where an MPLS domain nodes ("PE-n" and "P-m") are connected via two Ethernet switches ("ETH-n").

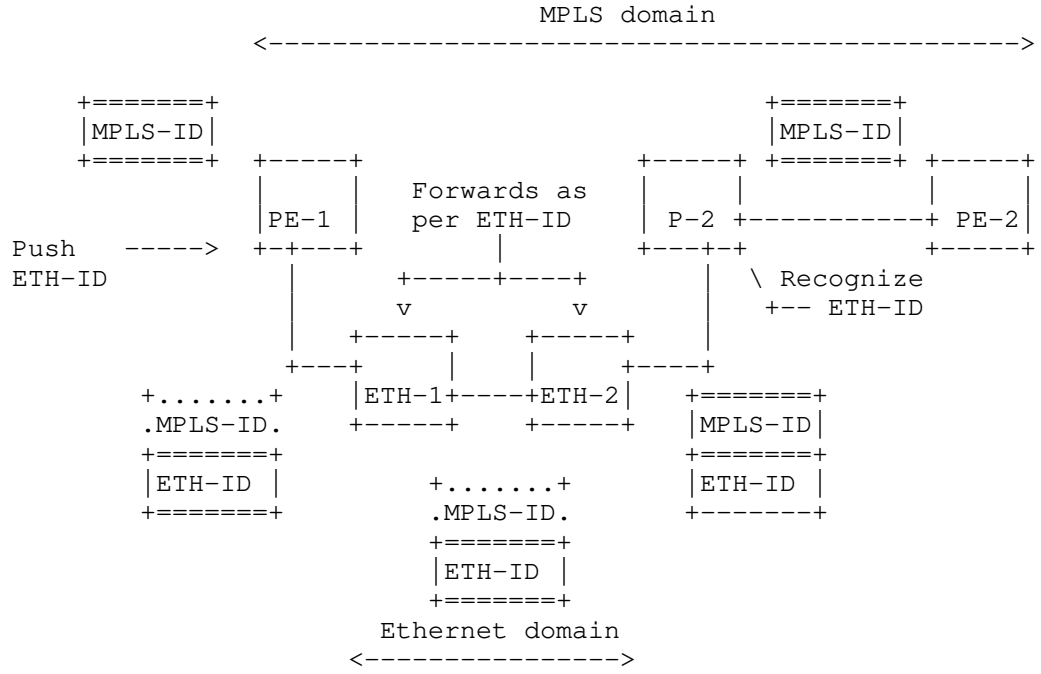



Figure 6: MPLS nodes interconnected by an Ethernet domain

"PE-1" uses the MPLS specific ID ("MPLS-ID"), but as it is connected to an Ethernet domain it has to push also an Ethernet-domain specific flow-ID ("VLAN+multicast-MAC", referred as "ETH-ID") before sending the packet to "ETH-1". The ethernet switch "ETH-1" can recognize the data flow based on the "ETH-ID" and it does forwarding towards "ETH-2". "ETH-2" switches the packet towards the MPLS node ("P-2"). "P-2" must be configured to receive the Ethernet Flow-ID specific multicast stream, but (as it is an MPLS node) it decodes the data flow ID based on the "MPLS-ID" fields of the received packet.

## 8. Summary

This document specifies a DetNet service model via related SAPs, Components/Segments and Terminology .

## 9. IANA Considerations

N/A.

## 10. Security Considerations

N/A.

## 11. Acknowledgements

The authors wish to thank Lou Berger, Norman Finn, Jouni Korhonen and the members of the data plane design team for their various contributions, comments and suggestions regarding this work.

## 12. Annex

This Annex contains some thoughts about scenarios where the service instance is shared by DetNet and regular traffic.

### 12.1. L2 service instance shared by regular and DetNet traffic

In case of a L2 VPN transport the service instance implements bridging. In MPLS based PSN there is a full mesh of PWs between service instances of PE nodes. Adding DetNet data flows to the network results in a somewhat modified PW structure, as a DetNet data flow requires its unique Flow-ID to be encoded in the packet.

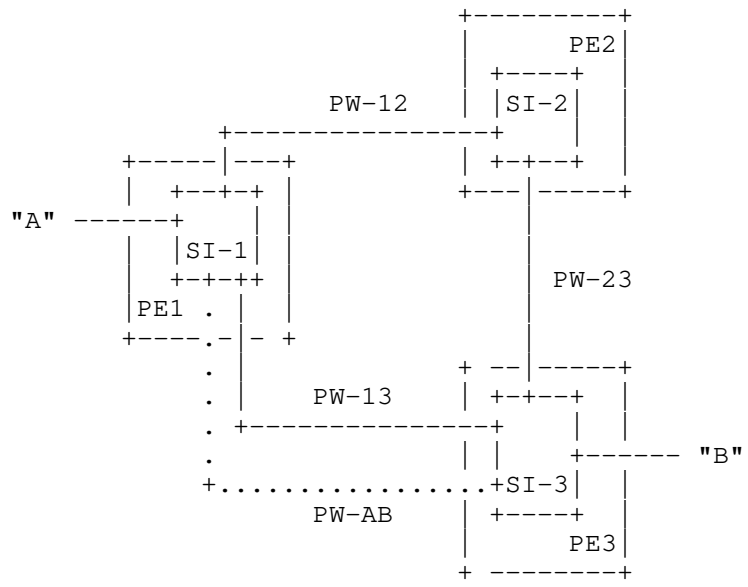



Figure 7: DetNet L2 VPN Service

Figure 7 shows a scenario where there is a DetNet data flow between the end-systems ("A" and "B"). "SI-n" denotes the L2 VPN service instance of "PEn". Regular traffic of the L2 VPN instance use "PW-12", "PW-13" and "PW-23". However for transport of DetNet traffic between "A" and "B" a separate PW ("PW-AB") have to be used. "PW-AB" is a somewhat special PW (called here "virtual PW") and it is treated differently than PWs used by regular traffic (i.e. PW-13, PW-12 and PW-23). Namely, "PW-AB" is used exclusivelly by the DetNet data flow between "A" and "B". "PW-AB" does not participate in flooding and no MAC addresses are associated with it (not considered for the MAC learning process). "PW-AB" may use the same LSP as "PW-13" or a dedicated one.

Regular traffic between "A" and "B" has an encapsulation [PW-13_label ; LSP_label], whereas DetNet data flow has [PW-AB_label ; LSP_label].

## 12.2. L3 service instance shared by regular and DetNet traffic

In case of a L3 DetNet service the service instance implements routing. In MPLS based PSN such a "routing service" can be provided by IP VPNs (rfc4364). However the IP VPN service add only a single label (VPN label) during forwarding, therefore the label stack does not contain a "control word" (i.e., there is no field to encode a

sequence number). So, transport of DetNet data flows requires the combination of IP VPN and PW technologies.

Adding DetNet data flows to the network results in a somewhat modified label stack structure, as a DetNet data flow requires its packet PW encapsulation (rfc6658).

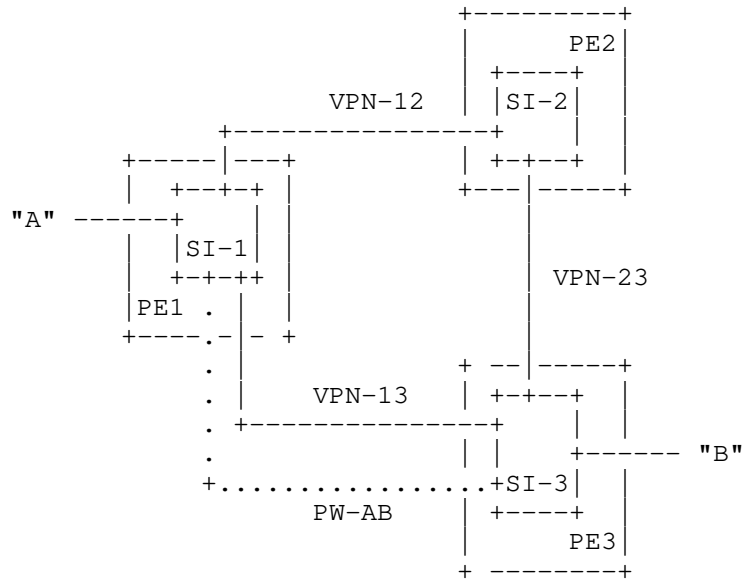



Figure 8: DetNet L3 VPN Service

Figure 8 shows a scenario where there is a DetNet data flow between the end-systems ("A" and "B"). "SI-n" denotes the L3 VPN service instance of "PEn". Regular traffic of the L3 VPN instance use as service label "VPN-12", "VPN-13" and "VPN-23". However for transport of DetNet traffic between "A" and "B" a PW ("PW-AB") have to be used, what ensures that DetNet data flow can be recognized by intermediate P nodes and a control world can be also present. "PW-AB" is used exclusivelly by the DetNet data flow between "A" and "B". "PW-AB" may use the same LSP as regular traffic (labeled by "VPN-13") or a dedicated one.

Regular traffic between "A" and "B" has an encapsulation [VPN-13_label ; LSP_label], whereas DetNet data flow has [PW-AB_label ; LSP_label].

# 13. References

## 13.1. Normative References

[draft-arch]
IETF, "Deterministic Networking Architecture", 2016, <https://datatracker.ietf.org/doc/draft-finn-detnet-architecture/>.

[draft-data-plane]
IETF, "DetNet Data Plane Protocol and Solution Alternatives", 2016, <https://datatracker.ietf.org/doc/draft-dt-detnet-dp-alt/>.

[I-D.ietf-detnet-use-cases]
Grossman, E., Gunther, C., Thubert, P., Wetterwald, P., Raymond, J., Korhonen, J., Kaneko, Y., Das, S., Zha, Y., Varga, B., Farkas, J., Goetz, F., and J. Schmitt, "Deterministic Networking Use Cases", draft-ietf-detnet-use-cases-09 (work in progress), March 2016.

[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>.

## 13.2. Informative References

[IETFDetNet]
IETF, "Charter for IETF DetNet Working Group", 2015, <https://datatracker.ietf.org/wg/detnet/charter/>.

# Authors' Addresses

Balazs Varga (editor)
Ericsson
Konyves Kalman krt. 11/B
Budapest 1097
Hungary

Email: balazs.a.varga@ericsson.com

Janos Farkas
Ericsson
Konyves Kalman krt. 11/B
Budapest 1097
Hungary

Email: janos.farkas@ericsson.com

Network Working Group
Internet Draft
Intended status: Informational
Expires: January 2017

Y. Zha
Y. Jiang
Huawei Technologies
L. Geng
China Mobile

July 8, 2016

# Deterministic Networking Flow Information Model
## draft-zha-detnet-flow-info-model-00

Abstract

Deterministic Networking (DetNet) provides end-to-end absolute delay and loss guarantee to serve real-time applications. DetNet is focused on a general approach that use techniques such as 1) data plane resources reservation for DetNet flows; 2) providing fixed path for DetNet flows; 3)sequentializng, replicating, and eliminating duplicate packets transmission [I-D.finn-detnet-architecture] to guarantee the worst case delay of DetNet flow while allow sharing among best effort traffic. Data flow information model is important to the DetNet work that it defines information be used by flow establishment and control protocols. This document describes and DetNet flow information model that represents the flow identifier, traffic description information so that can make resource reservation and provide differentiate service.

Table of Contents



## 1. Introduction

Deterministic service with both assured delay and data loss is promising to service providers. Due to lack of deterministic service provisioning mechanism there is no guarantee when deploying a time critical service [RFC3393]. Deterministic Networking (DetNet) tries to provide a solution to this issue with limited scope that the data flows are constrained with some maximum data rate properties. DetNet delivers assured end-to-end latency and packet loss by dedicating network resources to DetNet

flows while unused reserved resource are still open to best effort traffic.

In order to reserve proper amount of network resource to serve the DetNet flow, the DetNet flow first needs to be described with such parameters that can be understood by the network. Secondly, current flow description and resource reservation are mainly focused on bandwidth which is basically a statistical concept during a relative long observation interval. And also, there are different type of use cases those requires deterministic networking services [I-D.ietf-detnet-use-cases].

Data plane techniques such as queuing, shaping, scheduling and preemption are configured in a standard way to guarantee deterministic forwarding behavior in the network device. The controller or control plane takes the description of the DetNet flow and then translates into data plane level configuration to serve the flow. This is the key of DetNet as to define how to describe DetNet flow and how to reserve network resource for it. The flow description should be focused on traffic characteristics of real time service with parameters that could be converted to device level configurations.

An information model defines concepts in a uniform way, enabling formal mapping processes to be developed to the information model to a set of data models. This simplifies the process of constructing software to automate the policy management process. It also simplifies the language generation process, though that is beyond the scope of this document.

This document describes an information model for representing DetNet flow with comment concept and parameters of a DetNet service that can be mapped into device level configurations.

## 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 [RFC2119]. In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying [RFC2119] significance.

## 3. DetNet Flow Information Model

According to current charter, DetNet information model is to identify the information needed for flow establishment and control

and be used by reservation protocols and data plane configuration. The work will be independent from the protocol(s) used to control the flows. The DetNet information model presented in this document defines some common concept of DetNet flows with information that can be used for flow identification, flow monitoring, performance management, reservation protocol, and data plane configuration. For example, deterministic properties of controlled latency, low packet loss, low packet delay variation, and high reliability. More information can be added in the future. And each part of the information model can be used individually by different network function or network entities. The DetNet information model only defines what kind of information is needed and how it could be used. Data repository, data definition language, query language, implementation language, and protocol should not be defined here. More specifically, the information model can be used by a data model for different scenarios. As defined in [RFC3198], data model is "A mapping of the contents of an information model into a form that is specific to a particular type of data store or repository."

In this document, DetNet information model contains three sets of information, flow identifier information, flow metering statistics, and traffic description. More information will be added in the future version to make DetNet fully functional.

### 3.1. Flow Identifier

Flow identifier is the first step of flow description as DetNet requires differentiate service so it needs to be identified by the data plane. The DetNet service is described at flow level so each flow could have unique flow identifier.

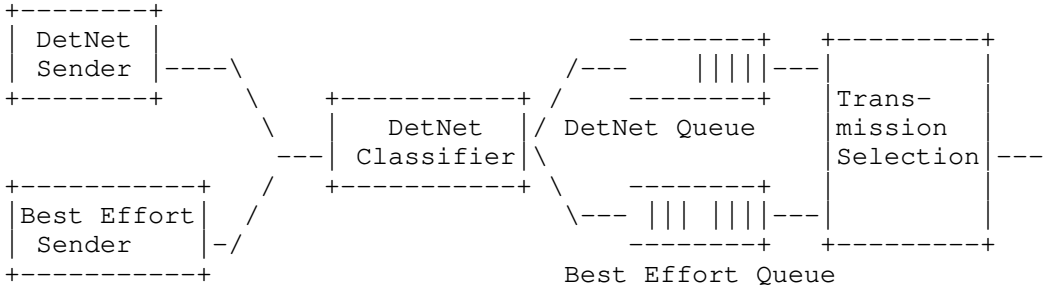



Figure 1. DetNet flow being classified

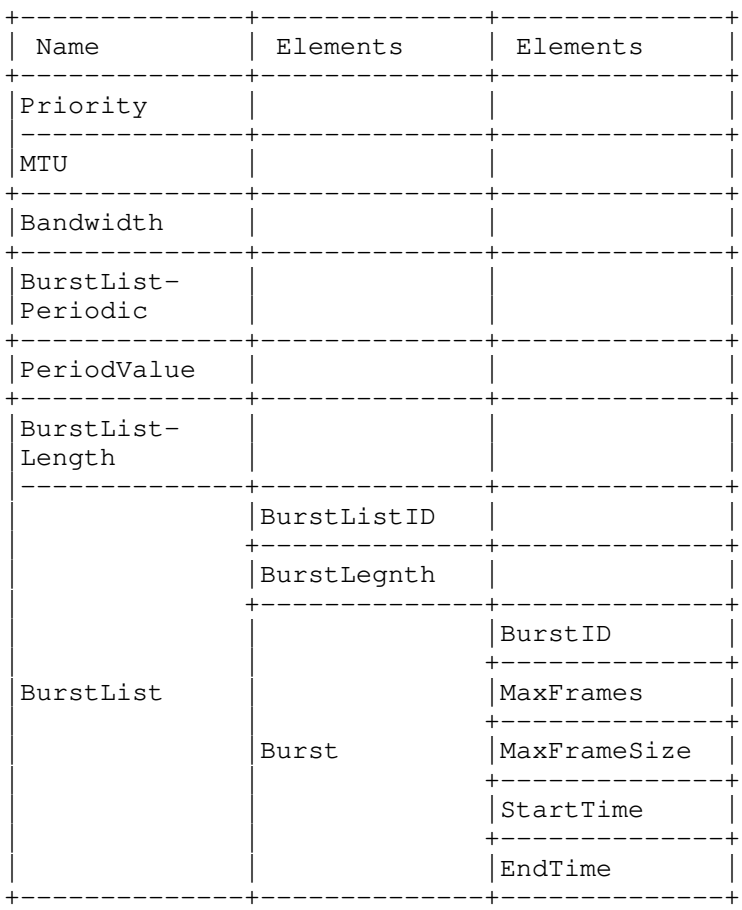
As shown in figure 1, network reserves dedicate resource for DetNet flows which will be identified first to use the resource. So a DetNet flow model should first contain information for flow identification. As shown in table 1, the information model has stream identifier and service type information for flow identification.

| Name | Elements |
|---|---|
| Stream Identifier | MAC Address |
| | StreamID |
| ServiceType | |

## 3.2. Flow Traffic Description

The information model should contain traffic description information to define the traffic profile from the source. Detnet flow defines the source guarantee that is the promise of source that the maximum amount traffic it can send. It is a kind of contract between the source and network who serve the flow. If the source is sending overload or different type of traffic, the overload or traffic does not match the predefined traffic profile will be not guaranteed. So the DetNet is that the source tells the network I will send the traffic like this, and the network will reserve the resource for the flow based on its traffic characteristics as defined in the model.

Unlike previous flow model or traffic profile which is mainly based on bandwidth of service, DetNet flow should be more accurate and at lower level for deterministic forwarding. For DetNet service provisioning which is focused on absolute worst case delay, the network needs to know not only the number of packets the flow will be sending but also when or during what period of time the source will be sending what amount of packets. Based on the architecture draft, there are two kinds of flows, synchronous one and asynchronous one. The information model of the DetNet flow with traffic description information is shown as below.




The basic idea is that the flow consists of a list of bursts. The Burst is a set of packets with burst duration. The burst is close related to service traffic pattern and also it is dependent on the data plane technique.

There are two basic requirements for traffic information definition, first it can be used to describe service; second the parameter defined here can be mapped to data plane configuration.

### 3.3. Flow Statistics

As a matter of fact, there is no mechanism to provide flow delay and loss parameter, which is also important for DetNet service. Keeping the knowledge of flow-based delay and loss information is also crucial for OAM and fault management.

The detail of flow metering statistic information in the information model will be proposed in next version.

| Name | Elements |
|---|---|
| MaxDelay | |
| MaxPacketLoss | |

## 4. Use of Flow Information Model

As defined in current charter, DetNet flow information model is used for flow establishment and control and can also be used by reservation protocols and YANG data models.

### 4.1. Data plane configuration

This section proposes a way to map information model parameters into network configuration. As defined in current charter, the DetNet data plane should be TSN compatible. Take TSN TAS (Time Aware Shaper) for example, the information defined in the flow model can be mapped to data plane parameter to configure TSN time aware shaper that provides a deterministic forwarding behavior for the flow.

As defined in previous section, information model contains traffic description of DetNet flow that can be used to configure data plane. In this section, take TSN time aware shaper as an example for data plane technique, mapping of data flow parameter to TAS configuration is presented.

The basic idea is that, the controller or network control plane takes data flow traffic description as the request and compute the associated time interval and control list of the TAS gate control function. Data flow model contains timing information of the DetNet flow as it arrives at T1 and ends at T2, which can be

mapped into control list of TAS to reserve an open gate for the DetNet flow for time period T1 to T2. As shown in figure 2, a DetNet flow with data traffic between T1 and T2 send the request to controller or control plane, and then the control plane uses the information to configure the TAS based on current status of TAS. Finally, the TAS function being configured with control list update for open gate transmission for this flow during T1 and T2. As a result, ideally, the flow will be transmitted immediately using the dedicated open gate time slot with absolute delay and loss guarantee.

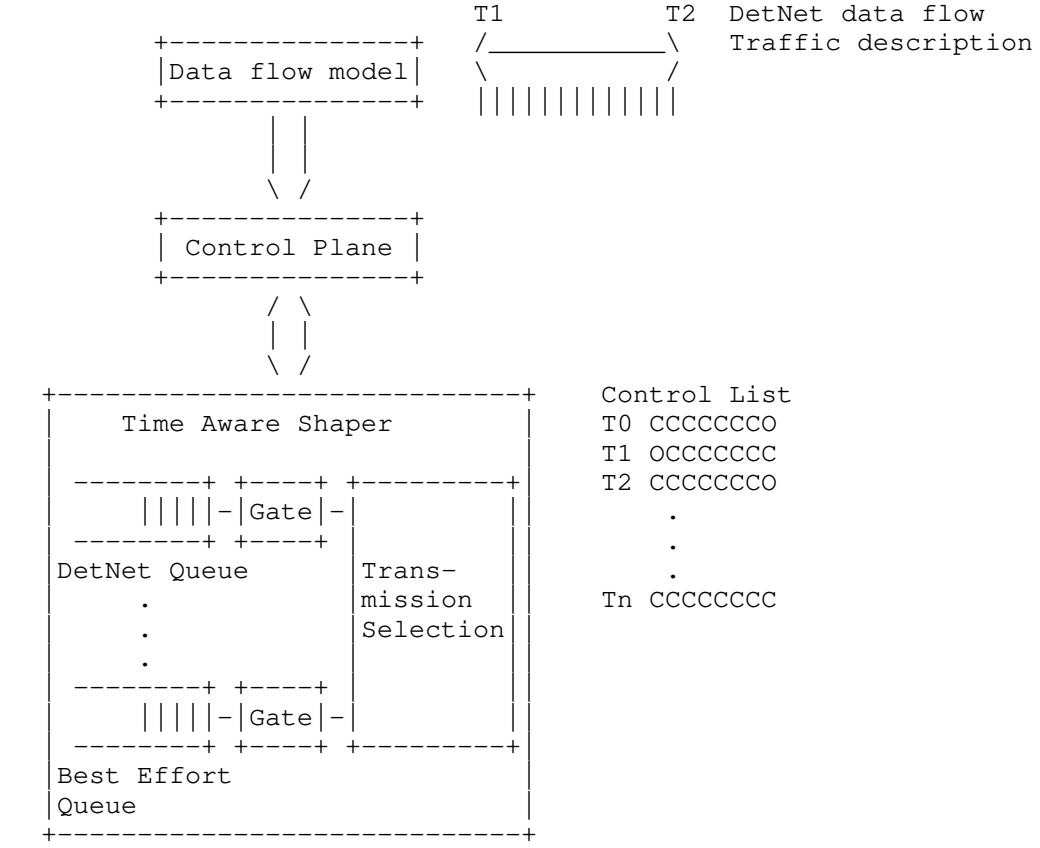



Figure 2. Mapping of Flow Model into TAS Configuration

## 5. Security Considerations

TBD

## 6. IANA Considerations

This document has no actions for IANA.

## 7. Acknowledgments

This document has benefited from reviews, suggestions, comments and proposed text provided by the following members, listed in alphabetical order: Jinchun Xu and Hengjun Zhu.

## 8. References

### 8.1. Normative References

[RFC2119] S. Bradner, "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3393] C. Demichelis, "IP Packet Delay Variation Metric for IP Performance Metrics (IPPM) ", RFC 3393, November 2002.

### 8.2. Informative References

[I-D.finn-detnet-problem-statement]

Finn, N. and P. Thubert, "Deterministic Networking Problem Statement", draft-finn-detnet-problem-statement-01 (work in progress), October 2014.

[I-D.finn-detnet-architecture]

Finn, N., Thubert, P., and M. Teener, "Deterministic Networking Architecture", draft-finn-detnet-architecture-01 (work in progress), March 2015.

Authors' Addresses

Yiyong Zha
Huawei Technologies
Email: zhayiyong@huawei.com

Yuanlong Jiang
Huawei Technologies
Email: jiangyuanlong@huawei.com

Liang Geng
China Mobile
Email: gengliang@chinamobile.com