DetNet
Internet-Draft
Intended status: Standards Track
Expires: February 19, 2017

N. Finn
P. Thubert
Cisco

August 18, 2016

Deterministic Networking Architecture
draft-finn-detnet-architecture-08

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 explicit routes for DetNet flows that do not rapidly change with the network topology; and 3) distributing data from DetNet flow packets over time and/or space to ensure delivery of each packet's data' in spite of the loss of a path. The capabilities can be managed by configuration, or by manual or automatic network management.

Table of Contents

1. Introduction

Deterministic Networking (DetNet) is a service that can be offered by a network to DetNet flows. DetNet provides these flows extremely low packet loss rates and assured maximum end-to-end delivery latency. 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. See [I-D.dt-detnet-dp-alt] for a discussion of specific techniques that can be used to identify DetNet Flows and assign them to specific paths through a network.

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 provided.

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 edge node
: An instance of a DetNet relay node that includes either a DetNet service layer proxy function for DetNet service protection (e.g. the addition or removal of packet sequencing information) for one or more end systems, or starts or terminates congestion protection at the DetNet transport layer, 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 node including a service layer function that interconnects different DetNet transport layer paths to provide service protection. A DetNet relay node can be a bridge, a router, a firewall, or any other system that participates in the DetNet service layer. It typically incorporates DetNet transport layer functions as well, in which case it is collocated with a transit node.

reservation
: A trail of configuration between source to destination(s) through transit nodes and subnets associated with a DetNet flow, to provide congestion protection.

DetNet service layer
: The layer at which service protection is provided, either packet sequencing, replication, and elimination (Section 3.4) or network coding (Section 3.5).

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. Optionally provides congestion protection over those paths. An MPLS LSR is an example of a DetNet transit node.

DetNet transport layer
: The layer that optionally provides congestion protection for DetNet flows over paths provided by the underlying network.

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:

- Congestion protection (Section 3.1).

- Explicit routes (Section 3.2).

- Service protection.

Congestion protection operates by reserving resources along the path of a DetNet Flow, e.g. buffer space or link bandwidth. Congestion protection greatly reduces, or even eliminates entirely, packet loss due to output packet congestion within the network, but it can only be supplied to a DetNet flow that is limited at the source to a maximum packet size and transmission rate.

Congestion protection addresses both of the DetNet QoS requirements (latency and packet loss). Given that DetNet nodes have a finite amount of buffer space, congestion protection 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. Service protection is the name for the mechanisms used by DetNet to address these losses. The mechanisms employed are constrained by the requirement to meet the users' latency requirements. Packet replication and elimination (Section 3.4) packet encoding Section 3.5 are described in this document to provide service protection; others may be found. Both mechanisms distribute the contents of DetNet flows over multiple paths in time and/or space, so that the loss of some of the paths does need not cause the loss of any packets. The paths are typically (but not necessarily) explicit routes, so that they cannot suffer temporary interruptions caused by the reconvergence of routing or bridging protocols.

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 plus service protection 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.

- Congestion protection 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. Congestion protection

The primary means by which DetNet achieves its QoS assurances is to reduce, or even 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 provide congestion protection. 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 for service protection 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. Service protection (Section 3.4 or Section 3.5) over explicit routes provides 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.

Jitter reduction is provided by the mechanisms described in Section 4.3 that also provide congestion protection.

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 spreading the data in a packet over multiple transmissions. One such method for service protection is described in this section, which sends the same packets over multiple paths. See also Section 3.5.

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:

- 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. Section 3.2.

- 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

- 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.

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

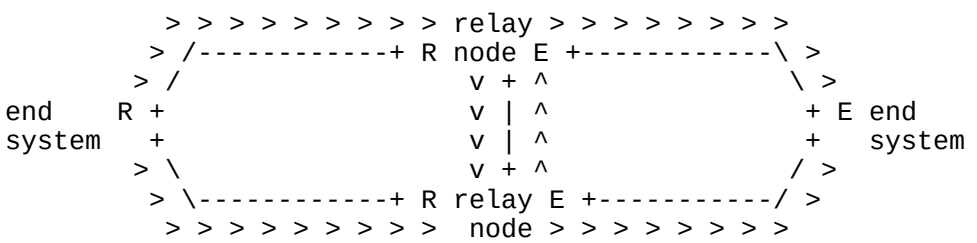


Figure 1

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.

If packet replication and elimination is used over paths providing congestion protection (Section 3.1), and member flows that take different-length paths through the network are combined, 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.

3.5. Packet encoding for service protection

There are methods for using multiple paths to provide service protection 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 service protection technique, assuming that the DetNet users'

needs for timeliness of delivery and freedom from interference with misbehaving DetNet flows can be met.

No specific mechanisms are defined here, at this time. This section will either be enhanced or removed. Contributions are invited.

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.

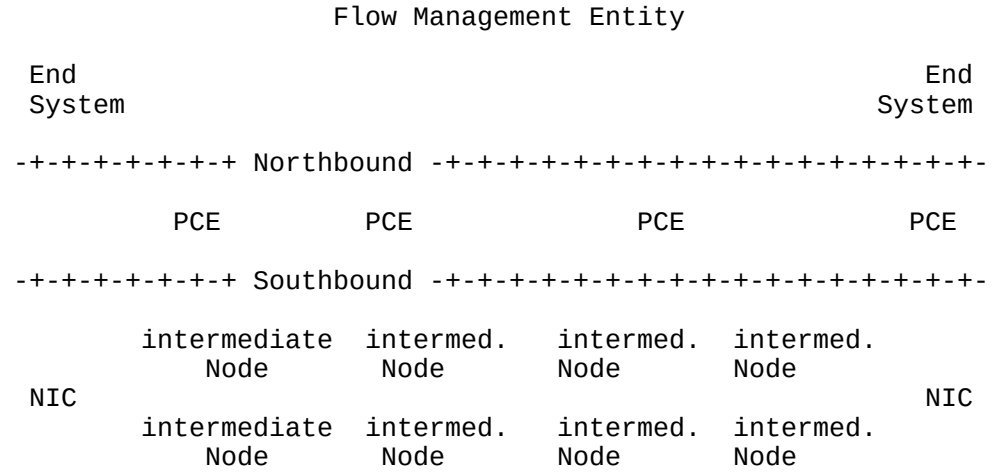


Figure 2

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

For the purposes of congestion protection, DetNet flows can be 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

DetNet achieves congestion protection and bounded delivery latency 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

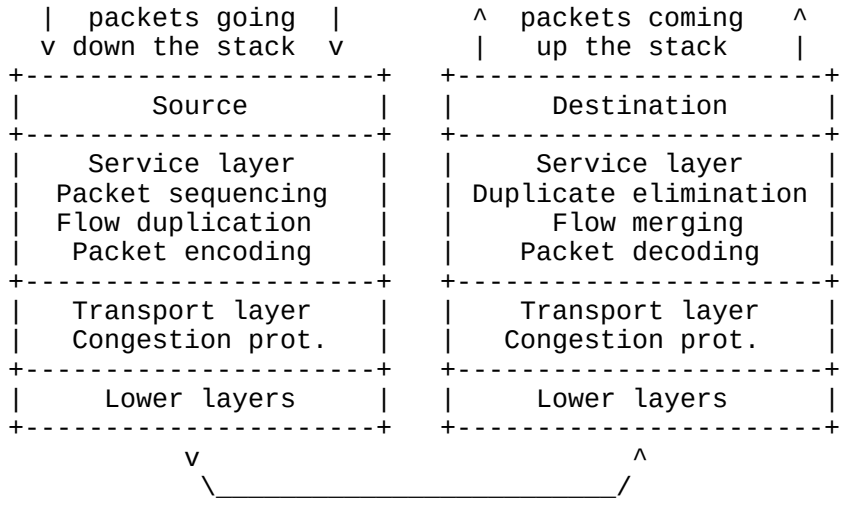


Figure 3

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. OAM is not shown in Figure 3; it 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 service protection, supplies the sequence number for packet replication and elimination (Section 3.4). 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.

Flow duplication
: As part of DetNet service protection, replicates packets 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 DetNet service protection, 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).

Packet encoding
: As part of DetNet service protection, as an alternative to packet sequencing and flow duplication, packet encoding combines the information in multiple DetNet packets, perhaps

from different DetNet compound flows, and transmits that information in packets on different DetNet member Flows. Peers with Packet decoding.

Packet decoding
: As part of DetNet service protection, as an alternative to flow merging and duplicate elimination, packet decoding takes packets from different DetNet member flows, and computes from those packets the original DetNet packets from the compound flows input to packet encoding. Peers with Packet encoding.

Congestio protection
: The DetNet transport layer provides congestion protection. See Section 4.3. The actual queuing and shaping mechaniss are typically provided by underlying subnet layers, but since these are can be closely associated with the means of providing paths for DetNet flows (e.g. MPLS LSPs or {VLAN, multicast destination MAC address} pairs), the path and the congestion protection are conflated in this figure.

Note that the packet sequencing and duplication elimination functions at the source and destination ends of a DetNet compound flow may be performed either in the end system or in a DetNet edge node. The reader must not confuse a DetNet edge function with other kinds of edge functions, e.g. an Label Edge Router, although the two functions may be performed together. The DetNet edge function is concerned with sequencing packets belonging to DetNet flows. The LER with encapsulating/decapsulating packets for transport, and is considered part of the network underlying the DetNet transport layer.

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 packet inspection. That is not to say that packet inspection to layer 4 or 5 addresses will not be used, or the capability standardized; but, there are alternatives.

A DetNet relay node can connect DetNet flows on different paths using different flow identification methods. 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.

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 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. 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).

6.2. 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 a DetNet 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 DetNet 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 DetNet edge node to initiate a per-to-peer protocol activity.)

6.3. 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.

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

- o the authentication and authorization of the controlling systems
- o 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, Lou Berger, and especially Michael Johas Teener, 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-03 (work in progress), August 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

detnet
Internet-Draft
Intended status: Standards Track
Expires: September 18, 2016

N. Finn
P. Thubert
Cisco

March 17, 2016

Deterministic Networking Problem Statement
draft-finn-detnet-problem-statement-05

Abstract

This paper documents the needs in various industries to establish multi-hop paths for characterized flows with deterministic properties .

Status of This Memo

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

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

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

This Internet-Draft will expire on September 18, 2016.

Table of Contents

1. Introduction

The Deterministic Networking Use Cases [I-D.grossman-detnet-use-cases] document illustrates that beyond the classical case of industrial automation and control systems (IACS), there are in fact multiple industries with strong and yet relatively similar needs for deterministic network services with latency guarantees and ultra-low packet loss.

The generalization of the needs for more deterministic networks have led to the IEEE 802.1 AVB Task Group becoming the Time-Sensitive Networking (TSN) [IEEE802.1TSNTG] Task Group (TG), with a much-expanded constituency from the industrial and vehicular markets.

Along with this expansion, the networks in consideration are becoming larger and structured, requiring deterministic forwarding beyond the LAN boundaries. For instance, IACS segregates the network along the broad lines of the Purdue Enterprise Reference Architecture (PERA) [ISA95], typically using deterministic local area networks for level 2 control systems, whereas public infrastructures such as Electricity Automation require deterministic properties over the Wide Area. The realization is now coming that the convergence of IT and Operational Technology (OT) networks requires Layer-3, as well as Layer-2, capabilities.

While the initial user base has focused almost entirely on Ethernet physical media and Ethernet-based bridging protocol (from several Standards Development Organizations), the need for Layer-3 expressed above, must not be confined to Ethernet and Ethernet-like media, and

while such media must be encompassed by any useful DetNet architecture, cooperation between IETF and other SDOs must not be limited to IEEE or IEEE 802. Furthermore, while the work completed and ongoing in other SDOs, and in IEEE 802 in particular, provide an obvious starting point for a DetNet architecture, we must not assume that these other SDOs' work confines the space in which the DetNet architecture progresses.

The properties of deterministic networks will have specific requirements for the use of routed networks to support these applications and a new model must be proposed to integrate determinism in IT technology. The proposed model should enable a fully scheduled operation orchestrated by a central controller, and may support a more distributed operation with probably lesser capabilities. In any fashion, the model should not compromise the ability of a network to keep carrying the sorts of traffic that is already carried today in conjunction with new, more deterministic flows.

Once the abstract model is agreed upon, the IETF will need to specify the signaling elements to be used to establish a path and the tagging elements to be used identify the flows that are to be forwarded along that path. The IETF will also need to specify the necessary protocols, or protocol additions, based on relevant IETF technologies, to implement the selected model.

As a result of this work, it will be possible to establish a multi-hop path over the IP network, for a particular flow with given timing and precise throughput requirements, and carry this particular flow along the multi-hop path with such characteristics as low latency and ultra-low jitter, duplication and elimination of packets over non-congruent paths for a higher delivery ratio, and/or zero congestion loss, regardless of the amount of other flows in the network.

Depending on the network capabilities and on the current state, requests to establish a path by an end-node or a network management entity may be granted or rejected, an existing path may be moved or removed, and DetNet flows exceeding their contract may face packet declassification and drop.

2. Terminology

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].

3. On Deterministic Networking

The Internet is not the only digital network that has grown dramatically over the last 30-40 years. Video and audio entertainment, and control systems for machinery, manufacturing processes, and vehicles are also ubiquitous, and are now based almost entirely on digital technologies. Over the past 10 years, engineers in these fields have come to realize that significant advantages in both cost and in the ability to accelerate growth can be obtained by basing all of these disparate digital technologies on packet networks.

The goals of Deterministic Networking are to enable the migration of applications that use special-purpose fieldbus technologies (HDMI, CANbus, ProfiBus, etc... even RS-232!) to packet technologies in general, and the Internet Protocol in particular, and to support both these new applications, and existing packet network applications, over the same physical network.

Considerable experience ([ODVA],[AVnu], [Profinet],[IEC62439], etc...) has shown that these applications need a some or all of a suite of features that includes:

1. Time synchronization of all host and network nodes (routers and/or bridges), accurate to something between 10 nanoseconds and 10 microseconds, depending on the application.

2. Support for critical packet flows that:

    * Can be unicast or multicast;

    * Need absolute guarantees of minimum and maximum latency end-to-end across the network; sometimes a tight jitter is required as well;

    * Need a packet loss ratio beyond the classical range for a particular medium, in the range of 1.0e-9 to 1.0e-12, or better, on Ethernet, and in the order of 1.0e-5 in Wireless Sensor mesh Networks;

    * Can, in total, absorb more than half of the network’s available bandwidth (that is, massive over-provisioning is ruled out as a solution);

    * Cannot suffer throttling, congestion feedback, or any other network-imposed transmission delay, although the flows can be meaningfully characterized either by a fixed, repeating

transmission schedule, or by a maximum bandwidth and packet size;

3. Multiple methods to schedule, shape, limit, and otherwise control the transmission of critical packets at each hop through the network data plane;

4. Robust defenses against misbehaving hosts, routers, or bridges, both in the data and control planes, with guarantees that a critical flow within its guaranteed resources cannot be affected by other flows whatever the pressures on the network;

5. One or more methods to reserve resources in bridges and routers to carry these flows.

Time synchronization techniques need not be addressed by an IETF Working Group; there are a number of standards available for this purpose, including IEEE 1588, IEEE 802.1AS, and more.

The multicast, latency, loss ratio, and non-throttling needs are made necessary by the algorithms employed by the applications. They are not simply the transliteration of fieldbus needs to a packet-based fieldbus simulation, but reflect fundamental mathematics of the control of a physical system.

With classical forwarding latency- and loss-sensitive packets across a network, interactions among different critical flows introduce fundamental uncertainties in delivery schedules. The details of the queuing, shaping, and scheduling algorithms employed by each bridge or router to control the output sequence on a given port affect the detailed makeup of the output stream, e.g. how finely a given flow’s packets are mixed among those of other flows.

This, in turn, has a strong effect on the buffer requirements, and hence the latency guarantees deliverable, by the next bridge or router along the path. For this reason, the IEEE 802.1 Time-Sensitive Networking Task Group has defined a new set of queuing, shaping, and scheduling algorithms that enable each bridge or router to compute the exact number of buffers to be allocated for each flow or class of flows.

Robustness is a common need for networking protocols, but plays a more important part in real-time control networks, where expensive equipment, and even lives, can be lost due to misbehaving equipment.

Reserving resources before packet transmission is the one fundamental shift in the behavior of network applications that is impossible to avoid. 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 un-throttled (though bandwidth limited) flows means that bridges and routers have to dedicate buffer resources to specific flows or to classes of flows. The requirements of each reservation have to be translated into the parameters that control each host's, bridge's, and router's queuing, shaping, and scheduling functions and delivered to the hosts, bridges, and routers.

4. Problem Statement

4.1. Supported topologies

In some use cases, the end point which run the application is involved in the deterministic networking operation, for instance by controlling certain aspects of its throughput such as rate or precise time of emission. In that case, the deterministic path is end-to-end from application host to application host.

On the other end, the deterministic portion of a path may be a tunnel between and ingress and an egress router. In any case, routers and switches in between should not need to be aware whether the path is end-to-end of a tunnel.

While it is clear that DetNet does not aim at setting up deterministic paths over the global Internet, there is still a lack of clarity on the limits of a domain where a deterministic path can be set up. These limits may depend in the technology that is used to seu th epath up, whether it is centralized or distributed.

4.2. Flow Characterization

Deterministic forwarding can only apply on flows with well-defined characteristics such as periodicity and burstiness. Before a path can be established to serve them, the expression of those characteristics, and how the network can serve them, for instance in shaping and forwarding operations, must be specified.

4.3. Centralized Path Computation and Installation

A centralized routing model, such as provided with a PCE, enables global and per-flow optimizations. 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 following classical Traffic Engineering (TE) models.

To enable a centralized model, DetNet should produce the complete SDN architecture with describes at a high level the interaction and data models to:

- report the topology and device capabilities to the central controller;
- establish a direct interface between the centralized PCE to each device under its control in order to enable a vertical signaling
- request a path setup for a new flow with particular characteristics over the service interface and control it through its life cycle;
- support for life cycle management for a path (instantiate/modify/update/delete)
- support for adaptability to cope with various events such as loss of a link, etc...
- expose the status of the path to the end devices (UNI interface)
- provide additional reliability through redundancy, in particular with packet replication and elimination;
- indicate the flows and packet sequences in-band with the flows;

4.4. Distributed Path Setup

Whether a distributed alternative without a PCE can be valuable could be studied as well. Such an alternative could for instance inherit from the Resource ReSerVation Protocol [RFC3209] (RSVP-TE) flows. But the focus of the work should be to deliver the centralized approach first.

To enable a RSVP-TE like functionality, the following steps would take place:

1. Neighbors and their capabilities are discovered and exposed to compute a path that fits the DetNet constraints, typically of latency, time precision and resource availability.

2. A constrained path is calculated with an improved version of CSPF that is aware of DetNet.

3. The path is installed using RSVP-TE, associated with flow identification, per-hop behavior such as replication and elimination, blocked resources, and flow timing information.

4. Traffic flows are transported through the MPLS-TE tunnel, using the reserved resources for this flow at each hop.

4.5. Duplicated data format

In some cases the duplication and elimination of packets over non-congruent paths is required to achieve a sufficiently high delivery ratio to meet application needs. In these cases, a small number of packet formats and supporting protocols are required (preferably, just one) to serialize the packets of a DetNet stream at one point in the network, replicate them at one or more points in the network, and discard duplicates at one or more other points in the network, including perhaps the destination host. Using an existing solution would be preferable to inventing a new one.

5. 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.

Typical control networks today rely on complete physical isolation to prevent rogue access to network resources. DetNet enables the virtualization of those networks over a converged IT/OT infrastructure. Doing so, DetNet introduces an additional risk that flows interact and interfere with one another as they share physical resources such as Ethernet trunks and radio spectrum. The requirement is that there is no possible data leak from and into a deterministic flow, and in a more general fashion there is no possible influence whatsoever from the outside on a deterministic flow. The expectation is that physical resources are effectively associated with a given flow at a given point of time. In that model, Time Sharing of physical resources becomes transparent to the individual flows which have no clue whether the resources are used by other flows at other times.

Security must cover:

- the protection of the signaling protocol
- the authentication and authorization of the controlling nodes
- the identification and shaping of the flows
- the isolation of flows from leakage and other influences from any activity sharing physical resources.

6. IANA Considerations

This document does not require an action from IANA.

7. Acknowledgments

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

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

8.2. 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 IEEE Audio Video Bridging (AVB) and Time-Sensitive Networking (TSN) standards.".

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

[EIP] http://www.odva.org/, "EtherNet/IP provides users with the network tools to deploy standard Ethernet technology (IEEE 802.3 combined with the TCP/IP Suite) for industrial automation applications while enabling Internet and enterprise connectivity data anytime, anywhere.", <http://www.odva.org/Portals/0/Library/Publications_Numbered/PUB00138R3_CIP_Adv_Tech_Series_EtherNetIP.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-03 (work in progress), March 2016.

[I-D.grossman-detnet-use-cases] Grossman, E., Gunther, C., Thubert, P., Wetterwald, P., Raymond, J., Korhonen, J., Kaneko, Y., Das, S., and Y. Zha, "Deterministic Networking Use Cases", draft-grossman-detnet-use-cases-01 (work in progress), November 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-09 (work in progress), November 2015.

[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-teas-yang-te-topo] Liu, X., Bryskin, I., Beeram, V., Saad, T., Shah, H., and O. Dios, "YANG Data Model for TE Topologies", draft-ietf-teas-yang-te-topo-02 (work in progress), October 2015.

[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.zhao-pce-pcep-extension-for-pce-controller]
Zhao, Q., Li, Z., Dhody, D., and C. Zhou, "PCEP Procedures and Protocol Extensions for Using PCE as a Central Controller (PCECC) of LSPs", draft-zhao-pce-pcep-extension-for-pce-controller-03 (work in progress), March 2016.

[IEC62439]
IEC, "Industrial communication networks - High availability automation networks - Part 3: Parallel Redundancy Protocol (PRP) and High-availability Seamless Redundancy (HSR) - IEC62439-3", 2012, <https://webstore.iec.ch/publication/7018>.

[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>.

[IEEE802.1Q-2011]
IEEE, "MAC Bridges and VLANs (IEEE 802.1Q-2011", 2011, <http://standards.ieee.org/getieee802/download/802.1Q-2011.pdf>.

[IEEE802.1Qat-2010]
IEEE, "Stream Reservation Protocol (IEEE 802.1Qat-2010)", 2010, <http://standards.ieee.org/getieee802/download/802.1Qat-2010.pdf>.

[IEEE802.1Qav]
IEEE, "Forwarding and Queuing (IEEE 802.1Qav-2009)", 2009, <http://standards.ieee.org/getieee802/download/802.1Qav-2009.pdf>.

[IEEE802.1TSNTG]
IEEE Standards Association, "IEEE 802.1 Time-Sensitive Networks Task Group", 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 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/>.

[MPLS] IETF, "Multiprotocol Label Switching", <https://datatracker.ietf.org/doc/charter-ietf-mpls/>.

[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/>.

[RFC2547] Rosen, E. and Y. Rekhter, "BGP/MPLS VPNs", RFC 2547, DOI 10.17487/RFC2547, March 1999, <http://www.rfc-editor.org/info/rfc2547>.

[RFC2702] Awduche, D., Malcolm, J., Agogbua, J., O'Dell, M., and J. McManus, "Requirements for Traffic Engineering Over MPLS", RFC 2702, DOI 10.17487/RFC2702, September 1999, <http://www.rfc-editor.org/info/rfc2702>.

[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>.

[RFC3272] Awduche, D., Chiu, A., Elwalid, A., Widjaja, I., and X. Xiao, "Overview and Principles of Internet Traffic Engineering", RFC 3272, DOI 10.17487/RFC3272, May 2002, <http://www.rfc-editor.org/info/rfc3272>.

[RFC3630] Katz, D., Kompella, K., and D. Yeung, "Traffic Engineering (TE) Extensions to OSPF Version 2", RFC 3630, DOI 10.17487/RFC3630, September 2003, <http://www.rfc-editor.org/info/rfc3630>.

[RFC3945] Mannie, E., Ed., "Generalized Multi-Protocol Label Switching (GMPLS) Architecture", RFC 3945, DOI 10.17487/RFC3945, October 2004, <http://www.rfc-editor.org/info/rfc3945>.

[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>.

[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>.

[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>.

[RFC5127] Chan, K., Babiarz, J., and F. Baker, "Aggregation of Diffserv Service Classes", RFC 5127, DOI 10.17487/RFC5127, February 2008, <http://www.rfc-editor.org/info/rfc5127>.

[RFC5151] Farrel, A., Ed., Ayyangar, A., and JP. Vasseur, "Inter-Domain MPLS and GMPLS Traffic Engineering -- Resource Reservation Protocol-Traffic Engineering (RSVP-TE) Extensions", RFC 5151, DOI 10.17487/RFC5151, February 2008, <http://www.rfc-editor.org/info/rfc5151>.

[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>.

[RFC5329] Ishiguro, K., Manral, V., Davey, A., and A. Lindem, Ed., "Traffic Engineering Extensions to OSPF Version 3", RFC 5329, DOI 10.17487/RFC5329, September 2008, <http://www.rfc-editor.org/info/rfc5329>.

[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>.

[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>.

[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>.

[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>.

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

[TiSCH] IETF, "IPv6 over the TSCH mode over 802.15.4", <https://datatracker.ietf.org/doc/charter-ietf-6tisch/>.

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

Authors' Addresses

Norm Finn
Cisco Systems
510 McCarthy Blvd
SJ-24
Milpitas, California 95035
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

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: May 12, 2016

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

November 9, 2015

Deterministic Networking Use Cases
draft-grossman-detnet-use-cases-01

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

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

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?

o 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 Use Cases

(This section was derived from draft-gunther-detnet-proaudio-req-01)

2.1. Introduction

The professional audio and video industry includes music and film content creation, broadcast, cinema, and live exposition as well as public address, media and emergency systems at large venues (airports, stadiums, churches, theme parks). These industries have already gone through the transition of audio and video signals from analog to digital, however the 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 attempting to transition to packet based infrastructure for distributing audio and video in order to reduce cost, increase routing flexibility, and integrate with existing IT infrastructure.

However, there are several requirements for making a network the primary infrastructure for audio and video which are not met by todays networks and these are our concern in this draft.

The principal requirement is that pro audio and video applications become able to establish streams that provide guaranteed (bounded) bandwidth and latency from the Layer 3 (IP) interface. Such streams can be created today within standards-based layer 2 islands however these are not sufficient to enable effective distribution over wider areas (for example broadcast events that span wide geographical areas).

Some proprietary systems have been created which enable deterministic streams at layer 3 however they are engineered networks in that they require careful configuration to operate, often require that the system be over designed, 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.

It would be highly desirable if such streams could be routed over the open Internet, however even intermediate solutions with more limited scope (such as enterprise networks) can provide a substantial improvement over todays networks, and a solution that only provides for the enterprise network scenario is an acceptable first step.

We also present more fine grained requirements of the audio and video industries such as safety and security, redundant paths, devices with limited computing resources on the network, and that reserved stream bandwidth is available for use by other best-effort traffic when that stream is not currently in use.

2.2. Fundamental Stream Requirements

The fundamental stream properties are guaranteed bandwidth and deterministic latency as described in this section. Additional stream requirements are described in a subsequent section.

2.2.1. Guaranteed Bandwidth

Transmitting audio and video streams is unlike common file transfer activities because guaranteed delivery 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 and stream playback is interrupted, which is unacceptable in for example a live concert. In some contexts large amounts of buffering can be used to provide enough delay to allow time for one or more retries, however this is not an effective solution when live interaction is involved, and is not considered an acceptable general solution for pro audio and video. (Have you ever tried speaking into a microphone through a sound system that has an echo coming back at you? It makes it almost impossible to speak clearly).

Providing a way to reserve a specific amount of bandwidth for a given stream is a key requirement.

2.2.2. Bounded and Consistent Latency

Latency in this context means the amount of time that passes between when a signal is sent over a stream and when it is received, for example the amount of time delay between when you speak into a microphone and when your voice emerges from the speaker. Any delay longer than about 10-15 milliseconds is noticeable by most live performers, and greater latency makes the system unusable because it

prevents them from playing in time with the other players (see slide 6 of [SRP_LATENCY]).

The 15ms latency bound is made even more challenging because it is often the case in network based music production with live electric instruments that multiple stages of signal processing are used, connected in series (i.e. from one to the other for example from guitar through a series of digital effects processors) in which case the latencies add, so the latencies of each individual stage must all together remain less than 15ms.

In some situations it is acceptable at the local location for content from the live remote site to be delayed to allow for a statistically acceptable amount of latency in order to reduce jitter. However, once the content begins playing in the local location any audio artifacts caused by the local network are unacceptable, especially in those situations where a live local performer is mixed into the feed from the remote location.

In addition to being bounded to within some predictable and acceptable amount of time (which may be 15 milliseconds or more or less depending on the application) the latency also has to be consistent. For example when playing a film consisting of a video stream and audio stream over a network, those two streams must be synchronized so that the voice and the picture match up. 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) buffer (delay) 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.

2.2.2.1. Optimizations

The 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].

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.

2.3. Additional Stream Requirements

The requirements in this section are more specific yet are common to multiple audio and video industry applications.

2.3.1. Deterministic Time to Establish Streaming

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 to another. Such systems currently use purpose-built hardware to switch feeds smoothly, however there is a current initiative in the broadcast industry to switch to a packet-based infrastructure (see [STUDIO_IP] and the ESPN DC2 use case described below).

2.3.2. 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 just reserve a ton of bandwidth and then never un-reserve it even though they are not using it, and soon they will have no bandwidth left.

2.3.3. Layer 3 Interconnecting Layer 2 Islands

As an intermediate step (short of providing guaranteed bandwidth across the open internet) it would be valuable to provide a way to connect multiple Layer 2 networks. For example layer 2 techniques could be used to create a LAN for a single broadcast studio, and several such studios could be interconnected via layer 3 links.

2.3.4. Secure Transmission

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.

2.3.5. Redundant Paths

On-air and other live media streams must be 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.

2.3.6. Link Aggregation

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 (see Bounded and Consistent Latency section above).

2.3.7. Traffic Segregation

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.

2.3.7.1. 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.

2.3.7.2. 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.

2.4. 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.

2.5. Security Considerations

Many industries that are moving from the point-to-point world to the digital network world have little understanding of the pitfalls that they can create for themselves with improperly implemented network infrastructure. DetNet should consider ways to provide security against DoS attacks in solutions directed at these markets. Some

considerations are given here as examples of ways that we can help new users avoid common pitfalls.

2.5.1. Denial of Service

One security pitfall that this author is aware of involves the use of technology that allows a presenter to throw the content from their tablet or smart phone onto the A/V system that is then viewed by all those in attendance. The facility introducing this technology was quite excited to allow such modern flexibility to those who came to speak. One thing they hadn’t realized was that since no security was put in place around this technology it left a hole in the system that allowed other attendees to "throw" their own content onto the A/V system.

2.5.2. Control Protocols

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). In addition, it would be desirable if the configuration protocols that are used to create the network paths used by the professional audio traffic could be designed to protect devices that are not meant to receive high-amplitude content from having such potentially damaging signals routed to them.

2.6. A State-of-the-Art Broadcast Installation Hits Technology Limits

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 details at [ESPN_DC2]).

In designing DC2 they replaced as much point-to-point technology as they possibly 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, and they were very happy with the results, however to interconnect these layer 2 LAN islands together they ended up using dedicated links because there is no standards-based routing solution available.

This is the kind of motivation we have to develop these standards because customers are ready and able to use them.

2.7. Acknowledgements

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

3. Utility Telecom Use Cases

(This section was derived from draft-wetterwald-detnet-utilities-reqs-02)

3.1. Overview

[I-D.finn-detnet-problem-statement] defines the characteristics of a deterministic flow as a data communication flow with a bounded latency, extraordinarily low frame loss, and a very narrow jitter. This document intends to define the utility requirements for deterministic networking.

Utility Telecom Networks

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.

Many utilities still rely on complex environments formed of multiple application-specific, proprietary networks. Information is siloed between operational areas. This prevents utility operations from realizing the operational efficiency benefits, visibility, and functional integration of operational information across grid applications and data networks. 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.

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. IPv6 is seen as the obvious future telecommunications technology for the Smart Grid. The Adhoc Group has disclosed, to the IEC coordination group, their conclusions at the end of 2014.

It is imperative that utilities participate in standards development bodies to influence the development of future solutions and to benefit from shared experiences of other utilities and vendors.

3.2. Telecommunications Trends and General telecommunications Requirements

These general telecommunications requirements are over and above the specific requirements of the use cases that have been addressed so far. These include both current and future telecommunications related requirements that should be factored into the network architecture and design.

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

3.2.1.1. 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.

3.2.2. Applications, Use cases and traffic patterns

Among the numerous applications and use cases that a utility deploys today, many rely on high availability and deterministic behaviour of the telecommunications networks. Protection use cases and generation control are the most demanding and can’t rely on a best effort approach.

3.2.2.1. Transmission use cases

Protection means not only the protection of the human operator but also the protection of the electric equipments and the preservation of the stability and frequency of the grid. If a default occurs on the transmission or the distribution of the electricity, important damages could occured to the human operator but also to very costly electrical equipments and perturb the grid leading to blackouts. The time and reliability requirements are very strong to avoid dramatic impacts to the electrical infrastructure.

3.2.2.1.1. Tele Protection

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.

- o 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.

- o 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 key elements that may impact Teleprotection performance include bandwidth rate of the Teleprotection system and its resiliency or failure recovery capacity. Transmission time, bandwidth utilization and resiliency are directly linked to the telecommunications equipments and the connections that are used to transfer the commands between relays.

3.2.2.1.1.1. Latency Budget Consideration

Delay requirements for utility networks may vary depending upon a number of parameters, such as the specific protection equipments used. 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.

3.2.2.1.1.2. Asymetric delay

In addition to minimal transmission delay, a differential protection telecommunications channel must be synchronous, i.e., experiencing symmetrical channel delay in transmit and receive paths. This requires special attention in jitter-prone packet networks. While optimally Teleprotection systems should support zero asymmetric delay, typical legacy relays can tolerate discrepancies of up to 750us.

The main tools available for lowering delay variation below this threshold are:

- A jitter buffer 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. This is the old TDM traditional way to fulfill this requirement.

- Traffic management tools ensure that the Teleprotection signals receive the highest transmission priority and minimize the number of jitter addition during the path. This is one way to meet the requirement in IP networks.

- Standard Packet-Based synchronization technologies, such as 1588-2008 Precision Time Protocol (PTP) and Synchronous Ethernet (Sync-E), can help maintain stable networks by keeping a highly accurate clock source on the different network devices involved.

3.2.2.1.1.2.1. Other traffic characteristics

- Redundancy: The existence in a system of more than one means of accomplishing a given function.

- Recovery time : The duration of time within which a business process must be restored after any type of disruption in order to avoid unacceptable consequences associated with a break in business continuity.

- performance management : In networking, a management function defined for controlling and analyzing different parameters/metrics such as the throughput, error rate.

- packet loss : One or more packets of data travelling across network fail to reach their destination.

3.2.2.1.1.2.2. Teleprotection network requirements

The following table captures the main network requirements (this is based on 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

3.2.2.1.2. Inter-Trip Protection scheme

Inter-tripping is the 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. The main use of such schemes is to ensure that protection at both ends of a faulted circuit will operate to isolate the equipment concerned. Inter-tripping schemes use signaling to convey a trip command to remote circuit breakers to isolate circuits.

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

3.2.2.1.3. Current Differential Protection Scheme

Current differential protection is commonly used for line protection, and is typical for protecting parallel circuits. A main advantage for differential protection is that, compared to overcurrent protection, it allows only the faulted circuit to be de-energized in case of a fault. At both end of the lines, the current is measured by the differential relays, and based on Kirchhoff’s law, 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. A fault in line 1 will cause overcurrent to be flowing in both lines, but because the current in line 2 is a through following current, this current is measured equal at both ends of the line, therefore the differential relays on line 2 will not trip line 2. Line 1 will be tripped, as the relays will not measure the same currents at both ends of the line. 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 requirements

3.2.2.1.4. Distance Protection Scheme

Distance (Impedance Relay) protection scheme is based on voltage and current measurements. A fault on a circuit will generally create a sag in the voltage level. If the ratio of voltage to current measured at the protection relay terminals, which equates to an impedance element, falls within a set threshold the circuit breaker will operate. The operating characteristics of this protection are based on the line characteristics. This means that when a fault appears on the line, the impedance setting in the relay is compared to the apparent impedance of the line from the relay terminals to the fault. If the relay setting is determined to be below the apparent impedance it is determined that the fault is within the zone of protection. When the transmission line length is under a minimum length, distance protection becomes more difficult to coordinate. In these instances the best choice of protection is 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

3.2.2.1.5. 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 merging unit 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

3.2.2.1.6. Intra-Substation Process Bus Communications

This use case describes the data flow from the CT/VT to the IEDs in the substation via the merging unit (MU). The CT/VT in the substation send the sampled value (analog voltage or current) to the Merging Unit (MU) over hard wire. The merging unit 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 MU through serial port, or IEEE 1588 protocol via 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

3.2.2.1.7. 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

3.2.2.1.8. 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. You will find herafter the table summarizing 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

3.2.2.2. Distribution use case

3.2.2.2.1. Fault Location Isolation and Service Restoration (FLISR)

As the name implies, 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. It is a self-healing feature whose purpose is to minimize the impact of faults by serving portions of the loads on the affected circuit by switching to other circuits. It reduces the number of customers that experience a sustained power outage by reconfiguring distribution circuits. This will likely be the first wide spread application of distributed intelligence in the grid. Secondary substations can be connected to multiple primary substations. Normally, static power switch statuses (open/closed) in the network dictate the power flow to secondary substations. Reconfiguring the network in the event of a fault is typically done manually on site to operate switchgear to energize/de-energize alternate paths. Automating the operation of substation switchgear allows the utility to have a more dynamic network where the flow of power can be altered under fault conditions but also during times of peak load. It allows the utility to shift peak loads around the network. Or, to be more precise, alters the configuration of the network to move loads

between different primary substations. The FLISR capability can be enabled in two modes:

- o Managed centrally from DMS (Distribution Management System), or
- o Executed locally through distributed control via intelligent switches and fault sensors.

There are 3 distinct sub-functions that are performed:

1. Fault Location Identification

This sub-function is initiated by SCADA inputs, such as lockouts, fault indications/location, and, also, by input from the Outage Management System (OMS), and in the future by inputs from fault-predicting devices. It determines the specific protective device, which has cleared the sustained fault, identifies the de-energized sections, and estimates the probable location of the actual or the expected fault. It distinguishes faults cleared by controllable protective devices from those cleared by fuses, and identifies momentary outages and inrush/cold load pick-up currents. This step is also referred to as Fault Detection Classification and Location (FDCL). This step helps to expedite the restoration of faulted sections through fast fault location identification and improved diagnostic information available for crew dispatch. Also provides visualization of fault information to design and implement a switching plan to isolate the fault.

2. Fault Type Determination

I. Indicates faults cleared by controllable protective devices by distinguishing between:

a. Faults cleared by fuses

b. Momentary outages

c. Inrush/cold load current

II. Determines the faulted sections based on SCADA fault indications and protection lockout signals

III. Increases the accuracy of the fault location estimation based on SCADA fault current measurements and real-time fault analysis

3. Fault Isolation and Service Restoration

Once the location and type of the fault has been pinpointed, the systems will attempt to isolate the fault and restore the non-faulted section of the network. This can have three modes of operation:

I. Closed-loop mode : This is initiated by the Fault location sub-function. It generates a switching order (i.e., sequence of switching) for the remotely controlled switching devices to isolate the faulted section, and restore service to the non-faulted sections. The switching order is automatically executed via SCADA.

II. Advisory mode : This is initiated by the Fault location sub-function. It generates a switching order for remotely and manually controlled switching devices to isolate the faulted section, and restore service to the non-faulted sections. The switching order is presented to operator for approval and execution.

III. Study mode : the operator initiates this function. It analyzes a saved case modified by the operator, and generates a switching order under the operating conditions specified by the operator.

With the increasing volume of data that are collected through fault sensors, utilities will use Big Data query and analysis tools to study outage information to anticipate and prevent outages by detecting failure patterns and their correlation with asset age, type, load profiles, time of day, weather conditions, and other conditions to discover conditions that lead to faults and take the necessary preventive and corrective measures.

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 9: FLISR Communication Requirements

3.2.2.3. Generation use case

3.2.2.3.1. Frequency Control / Automatic Generation Control (AGC)

The system frequency should be maintained within a very narrow band. Deviations from the acceptable frequency range are detected and forwarded to the Load Frequency Control (LFC) system so that required up or down generation increase / decrease pulses can be sent to the power plants for frequency regulation. The trend in system frequency is a measure of mismatch between demand and generation, and is a necessary parameter for load control in interconnected systems.

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. Since a power grid requires that generation and load closely balance moment by moment, frequent adjustments to the output of generators are necessary. The balance can be judged by measuring the system frequency; if it is increasing, more power is being generated than used, and all machines in the system are accelerating. If the system frequency is decreasing, more demand is on the system than the instantaneous generation can provide, and all generators are slowing down.

Where the grid has tie lines to adjacent control areas, automatic generation control helps maintain the power interchanges over the tie lines at the scheduled levels. The AGC takes into account various parameters including the most economical units to adjust, the coordination of thermal, hydroelectric, and other generation types, and even constraints related to the stability of the system and capacity of interconnections to other power grids.

For the purpose of AGC we use static frequency measurements and averaging methods are used to get a more precise measure of system frequency in steady-state conditions.

During disturbances, more real-time dynamic measurements of system frequency are taken using PMUs, especially when different areas of the system exhibit different frequencies. But that is outside the scope of this use case.

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 10: FCAG Communication Requirements

3.2.3. 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 SPS 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.

Due to vast distances between transmission substations (for example as far as 280km apart), the fiber signal can be amplified to reach a distance of 280 km without attenuation.

3.2.4. 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 (IEEE 1588), distributing the synchronization signal over the IP/MPLS network.

The Precision Time Protocol (PTP) is defined in IEEE standard 1588. PTP is applicable to distributed systems consisting of one or more nodes, communicating over a network. Nodes are modeled as containing a real-time clock that may be used by applications within the node for various purposes such as generating time-stamps for data or ordering events managed by the node. The protocol 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 eliminates cyclic forwarding of PTP messages between communication paths.

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. 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 asymmetry in the paths taken by event messages. Asymmetry is not detectable by PTP, however, if known, PTP corrects for asymmetry.

A time-stamp event is generated at the time of transmission and reception of any event message. The time-stamp event occurs when the message's timestamp point crosses the boundary between the node and the network.

IEC 61850 will recommend the use of the IEEE PTP 1588 Utility Profile (as defined in IEC 62439-3 Annex B) which offers the support of redundant attachment of clocks to Paralell Redundancy Protcol (PRP) and High-availability Seamless Redundancy (HSR) networks.

3.3. IANA Considerations

This memo includes no request to IANA.

3.4. Security Considerations

3.4.1. Current Practices and Their Limitations

Grid monitoring and control devices are already targets for cyber attacks and legacy telecommunications protocols have many intrinsic network related vulnerabilities. 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).

- Bump in the wire (BITW) solutions : A hardware device is added to provide IPSec services between two routers that are not capable of IPSec functions. This special IPsec device will intercept then intercept outgoing datagrams, add IPSec protection to them, and strip it off incoming datagrams. BITW can all IPSec to legacy hosts and can retrofit non-IPSec routers to provide security benefits. The disadvantages are complexity and cost.

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.

3.4.2. Security Trends in Utility Networks

Although advanced telecommunications networks can assist in transforming the energy industry, 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’s 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:

1. IP enables a rich set of features and capabilities to enhance the security posture

2. 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.

3.5. Acknowledgements

Faramarz Maghsoodlou, Ph. D. IoT Connected Industries and Energy Practice Cisco

Pascal Thubert, CTAO Cisco

4. Building Automation Systems Use Cases

4.1. Introduction

Building Automation System (BAS) is a system that manages various equipment and sensors in buildings (e.g., heating, cooling and ventilating) for improving residents' comfort, reduction of energy consumption and automatic responses in case of failure and emergency. For example, BAS measures temperature of a room by using various sensors and then controls the HVAC (Heating, Ventilating, and air Conditioning) system automatically to maintain the temperature level and minimize the energy consumption.

There are typically two layers of network in a BAS. Upper one is called management network and the lower one is called field network. In management networks, an IP-based communication protocol is used while in field network, non-IP based communication protocols (a.k.a., field protocol) are mainly used.

There are many field protocols used in today's deployment in which some medium access control and physical layers protocols are standards-based and others are proprietary based. Therefore the BAS needs to have multiple MAC/PHY modules and interfaces to make use of multiple field protocols based devices. This situation not only makes BAS more expensive with large development cycle of multiple devices but also creates the issue of vendor lock-in with multiple types of management applications.

The other issue with some of the existing field networks and protocols are security. When these protocols and network were developed, it was assumed that the field networks are isolated physically from external networks and therefore the network and protocol security was not a concern. However, in today's world many BASes are managed remotely and is connected to shared IP networks and it is also not uncommon that same IT infrastructure is used be it office, home or in enterprise networks. Adding network and protocol security to existing system is a non-trivial task.

This document first describes the BAS functionalities, its architecture and current deployment models. Then we discuss the use cases and field network requirements that need to be satisfied by deterministic networking.

4.2. BAS Functionality

Building Automation System (BAS) is a system that manages various devices in buildings automatically. BAS primarily performs the following functions:

- Measures states of devices in a regular interval. For example, temperature or humidity or illuminance of rooms, on/off state of room lights, open/close state of doors, FAN speed, valve, running mode of HVAC, and its power consumption.

- Stores the measured data into a database (Note: The database keeps the data for several years).

- Provides the measured data for BAS operators for visualization.

- Generates alarms for abnormal state of devices (e.g., calling operator's cellular phone, sending an e-mail to operators and so on).

- Controls devices on demand.

- Controls devices with a pre-defined operation schedule (e.g., turn off room lights at 10:00 PM).

4.3. BAS Architecture

A typical BAS architecture is described below in Figure 1. There are several elements in a BAS.

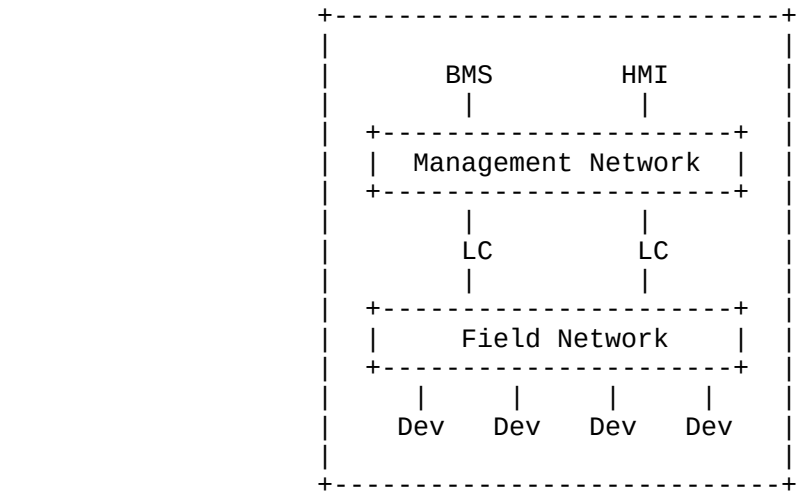


Figure 1: BAS architecture

Human Machine Interface (HMI): It is commonly a computing platform (e.g., desktop PC) used by operators. Operators perform the following operations through HMI.

- Monitoring devices: HMI displays measured device states. For example, latest device states, a history chart of states, a popup window with an alert message. Typically, the measured device states are stored in BMS (Building Management Server).

- Controlling devices: HMI provides ability to control the devices. For example, turn on a room light, set a target temperature to HVAC. Several parameters (a target device, a control value, etc.), can be set by the operators which then HMI sends to a LC (Local Controller) via the management network.

- Configuring an operational schedule: HMI provides scheduling capability through which operational schedule is defined. For example, schedule includes 1) a time to control, 2) a target device to control, and 3) a control value. A specific operational example could be turn off all room lights in the building at 10:00 PM. This schedule is typically stored in BMS.

Building Management Server (BMS) collects device states from LCs (Local Controllers) and stores it into a database. According to its configuration, BMS executes the following operation automatically.

- BMS collects device states from LCs in a regular interval and then stores the information into a database.

- BMS sends control values to LCs according to a pre-configured schedule.

- BMS sends an alarm signal to operators if it detects abnormal devices states. For example, turning on a red lamp, calling operators' cellular phone, sending an e-mail to operators.

BMS and HMI communicate with Local Controllers (LCs) via IP-based communication protocol standardized by BACnet/IP [bacnetip], KNX/IP [knx]. These protocols are commonly called as management protocols. LCs measure device states and provide the information to BMS or HMI. These devices may include HVAC, FAN, doors, valves, lights, sensors (e.g., temperature, humidity, and illuminance). LC can also set control values to the devices. LC sometimes has additional functions, for example, sending a device state to BMS or HMI if the device state exceeds a certain threshold value, feedback control to a device to keep the device state at a certain state. Typical example of LC is a PLC (Programmable Logic Controller).

Each LC is connected with a different field network and communicates with several tens or hundreds of devices via the field network. Today there are many field protocols used in the field network. Based on the type of field protocol used, LC interfaces and its hardware/software could be different. Field protocols are currently non-IP based in which some of them are standards-based (e.g., LonTalk [lontalk], Modbus [modbus], Profibus [profibus], FL-net [flnet],) and others are proprietary.

4.4. Deployment Model

An example BAS system deployment model for medium and large buildings is depicted in Figure 2 below. In this case the physical layout of the entire system spans across multiple floors in which there is normally 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.

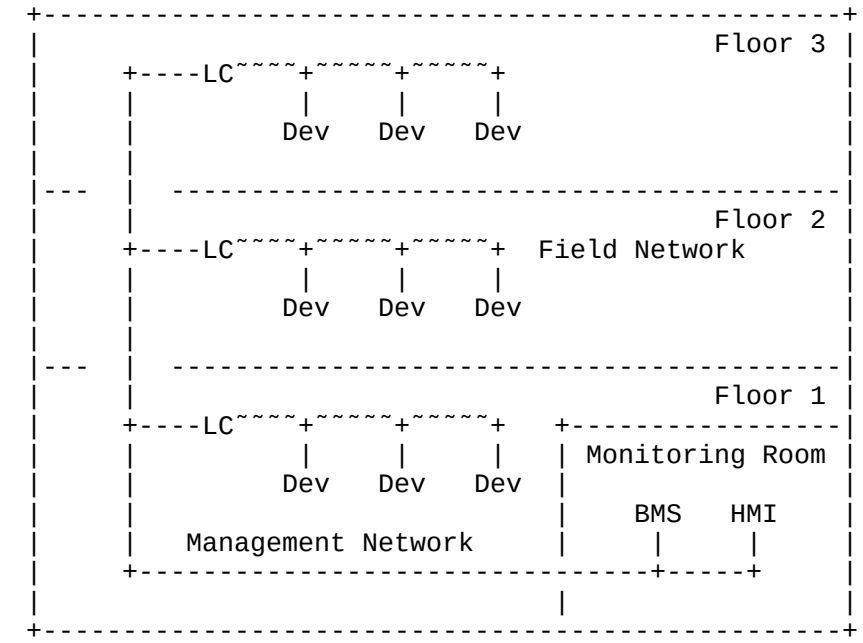


Figure 2: Deployment model for Medium/Large Buildings

Each LC is then connected to the monitoring room via the management network. In this scenario, the management functions are performed locally and reside within the building. In most cases, fast Ethernet (e.g. 100BASE-TX) is used for the management network. In the field network, variety of physical interfaces such as RS232C, and RS485 are

used. Since management network is non-real time, Ethernet without quality of service is sufficient for today's deployment. However, the requirements are different for field networks when they are replaced by either Ethernet or any wireless technologies supporting real time requirements (Section 3.4).

Figure 3 depicts a deployment model in which the management can be hosted remotely. This deployment is becoming popular for small office and residential buildings whereby having a standalone monitoring system is not a cost effective solution. In such scenario, multiple buildings are managed by a remote management monitoring system.

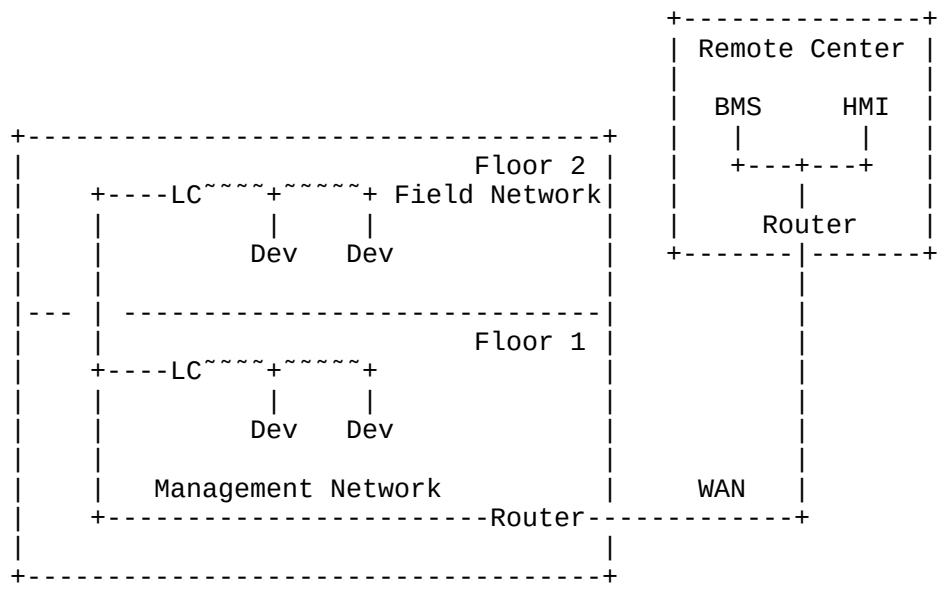


Figure 3: Deployment model for Small Buildings

In either case, interoperability today is only limited to the management network and its protocols. In existing deployment, there are limited interoperability opportunity in the field network due to its nature of non-IP-based design and requirements.

4.5. Use cases and Field Network Requirements

In this section, we describe several use cases and corresponding network requirements.

4.5.1. Environmental Monitoring

In this use case, LCs measure environmental data (e.g. temperatures, humidity, illuminance, CO2, etc.) from several sensor devices at each measurement interval. LCs keep latest value of each sensor. BMS sends data requests to LCs to collect the latest values, then stores the collected values into a database. Operators check the latest environmental data that are displayed by the HMI. BMS also checks the collected data automatically to notify the operators if a room condition was going to bad (e.g., too hot or cold). The following table lists the field network requirements in which the number of devices in a typical building will be ~100s per LC.

Metric	Requirement
Measurement interval	100 msec
Availability	99.999 %

Table 11: Field Network Requirements for Environmental Monitoring

There is a case that BMS sends data requests at each 1 second in order to draw a historical chart of 1 second granularity. Therefore 100 msec measurement interval is sufficient for this use case, because typically 10 times granularity (compared with the interval of data requests) is considered enough accuracy in this use case. A LC needs to measure values of all sensors connected with itself at each measurement interval. Each communication delay in this scenario is not so critical. The important requirement is completing measurements of all sensor values in the specified measurement interval. The availability in this use case is very high (Three 9s).

4.5.2. Fire Detection

In the case of fire detection, HMI needs to show a popup window with an alert message within a few seconds after an abnormal state is detected. BMS needs to do some operations if it detects fire. For example, stopping a HVAC, closing fire shutters, and turning on fire sprinklers. The following table describes requirements in which the number of devices in a typical building will be ~10s per LC.

Metric	Requirement
Measurement interval	10s of msec
Communication delay	< 10s of msec
Availability	99.9999 %

Table 12: Field Network Requirements for Fire Detection

In order to perform the above operation within a few seconds (1 or 2 seconds) after detecting fire, LCs should measure sensor values at a regular interval of less than 10s of msec. If a LC detects an abnormal sensor value, it sends an alarm information to BMS and HMI immediately. BMS then controls HVAC or fire shutters or fire sprinklers. HMI then displays a pop up window and generates the alert message. Since the management network does not operate in real time, and software run on BMS or HMI requires 100s of ms, the communication delay should be less than ~10s of msec. The availability in this use case is very high (Four 9s).

4.5.3. Feedback Control

Feedback control is used to keep a device state at a certain value. For example, keeping a room temperature at 27 degree Celsius, keeping a water flow rate at 100 L/m and so on. The target device state is normally pre-defined in LCs or provided from BMS or from HMI.

In feedback control procedure, a LC repeats the following actions at a regular interval (feedback interval).

1. The LC measures device states of the target device.

2. The LC calculates a control value by considering the measured device state.

3. The LC sends the control value to the target device.

The feedback interval highly depends on the characteristics of the device and a target quality of control value. While several tens of milliseconds feedback interval is sufficient to control a valve that regulates a water flow, controlling DC motors requires several milliseconds interval. The following table describes the field network requirements in which the number of devices in a typical building will be ~10s per LC.

Metric	Requirement
Feedback interval	~10ms - 100ms
Communication delay	< 10s of msec
Communication jitter	< 1 msec
Availability	99.9999 %

Table 13: Field Network Requirements for Feedback Control

Small communication delay and jitter are required in this use case in order to provide high quality of feedback control. This is currently offered in production environment with hgh availability (Four 9s).

4.6. Security Considerations

Both network and physical security of BAS are important. While physical security is present in today's deployment, adequate network security and access control are either not implemented or configured properly. This was sufficient in networks while they are isolated and not connected to the IT or other infrastructure networks but when IT and OT (Operational Technology) are connected in the same infrastructure network, network security is essential. The management network being an IP-based network does have the protocols and knobs to enable the network security but in many cases BAS for example, does not use device authentication or encryption for data in transit. On the contrary, many of today's field networks do not provide any security at all. Following are the high level security requirements that the network should provide:

- 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
- Availability of network data for normal and disaster scenario

5. Wireless for Industrial Use Cases

(This section was derived from draft-thubert-6tisch-4detnet-01)

5.1. Introduction

The emergence of wireless technology has enabled a variety of new devices to get interconnected, at a very low marginal cost per device, at any distance ranging from Near Field to interplanetary, and in circumstances where wiring may not be practical, for instance on fast-moving or rotating devices.

At the same time, a new breed of Time Sensitive Networks is being developed to enable traffic that is highly sensitive to jitter, quite sensitive to latency, and with a high degree of operational criticality so that loss should be minimized at all times. Such traffic is not limited to professional Audio/ Video networks, but is also found in command and control operations such as industrial automation and vehicular sensors and actuators.

At IEEE802.1, the Audio/Video Task Group [IEEE802.1TSNTG] Time Sensitive Networking (TSN) to address Deterministic Ethernet. The Medium access Control (MAC) of IEEE802.15.4 [IEEE802154] has evolved with the new TimeSlotted Channel Hopping (TSCH) [RFC7554] mode for deterministic industrial-type applications. TSCH was introduced with the IEEE802.15.4e [IEEE802154e] amendment and will be wrapped up in the next revision of the IEEE802.15.4 standard. For all practical purpose, this document is expected to be insensitive to the future versions of the IEEE802.15.4 standard, which is thus referenced undated.

Though at a different time scale, both TSN and TSCH standards provide Deterministic capabilities to the point that a packet that pertains to a certain flow crosses the network from node to node following a very precise schedule, as a train that leaves intermediate stations at precise times along its path. With TSCH, time is formatted into timeSlots, and an individual cell is allocated to unicast or broadcast communication at the MAC level. The time-slotted operation reduces collisions, saves energy, and enables to more closely engineer the network for deterministic properties. The channel hopping aspect is a simple and efficient technique to combat multi-path fading and co-channel interferences (for example by Wi-Fi emitters).

The 6TiSCH Architecture [I-D.ietf-6tisch-architecture] defines a remote monitoring and scheduling management of a TSCH network by a Path Computation Element (PCE), which cooperates with an abstract Network Management Entity (NME) to manage timeSlots and device

resources in a manner that minimizes the interaction with and the load placed on the constrained devices.

This Architecture applies the concepts of Deterministic Networking on a TSCH network to enable the switching of timeSlots in a G-MPLS manner. This document details the dependencies that 6TiSCH has on PCE [PCE] and DetNet [I-D.finn-detnet-architecture] to provide the necessary capabilities that may be specific to such networks. In turn, DetNet is expected to integrate and maintain consistency with the work that has taken place and is continuing at IEEE802.1TSN and AVnu.

5.2. Terminology

Readers are expected to be familiar with all the terms and concepts that are discussed in "Multi-link Subnet Support in IPv6" [I-D.ietf-ipv6-multilink-subnets].

The draft uses terminology defined or referenced in [I-D.ietf-6tisch-terminology] and [I-D.ietf-roll-rpl-industrial-applicability].

The draft also conforms to the terms and models described in [RFC3444] and uses the vocabulary and the concepts defined in [RFC4291] for the IPv6 Architecture.

5.3. 6TiSCH Overview

The scope of the present work is a subnet that, in its basic configuration, is made of a TSCH [RFC7554] MAC Low Power Lossy Network (LLN).

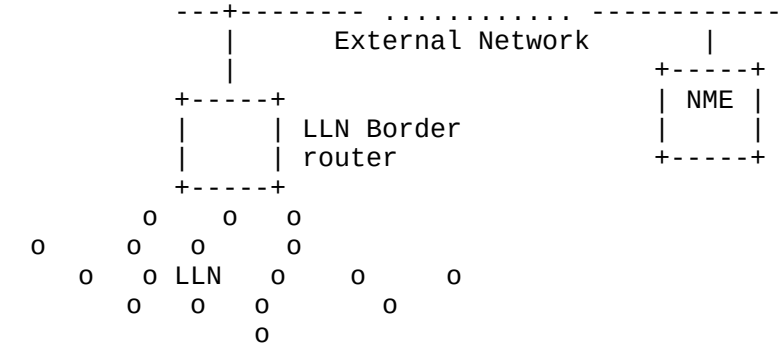


Figure 4: Basic Configuration of a 6TiSCH Network

In the extended configuration, a Backbone Router (6BBR) federates multiple 6TiSCH in a single subnet over a backbone. 6TiSCH 6BBRs synchronize with one another over the backbone, so as to ensure that the multiple LLNs that form the IPv6 subnet stay tightly synchronized.

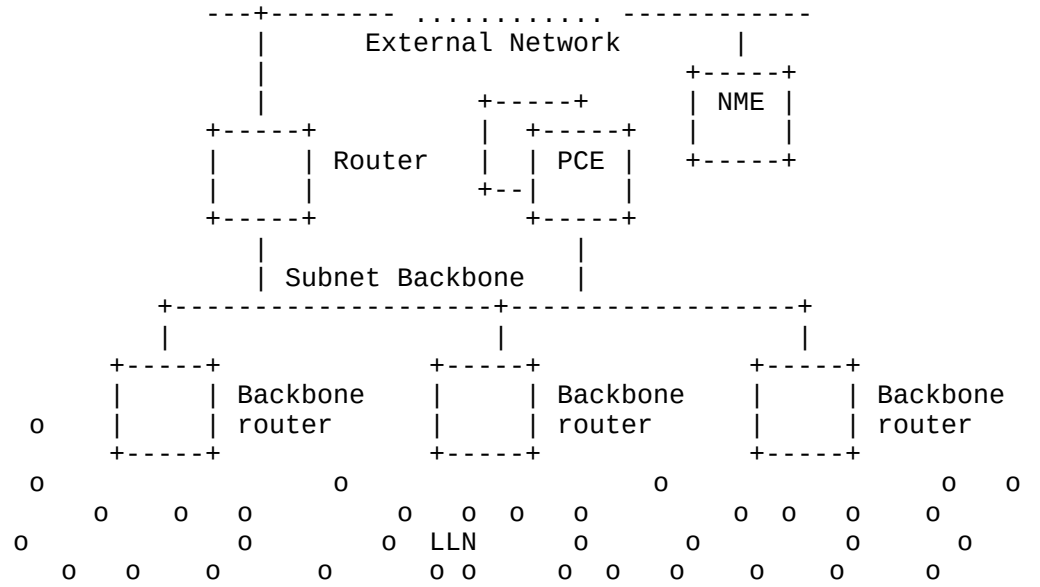


Figure 5: Extended Configuration of a 6TiSCH Network

If the Backbone is Deterministic, then the Backbone Router ensures that the end-to-end deterministic behavior is maintained between the LLN and the backbone. This SHOULD be done in conformance to the DetNet Architecture [I-D.finn-detnet-architecture] which studies Layer-3 aspects of Deterministic Networks, and covers networks that span multiple Layer-2 domains. One particular requirement is that the PCE MUST be able to compute a deterministic path and to end across the TSCH network and an IEEE802.1 TSN Ethernet backbone, and DetNet MUST enable end-to-end deterministic forwarding.

6TiSCH defines the concept of a Track, which is a complex form of a uni-directional Circuit ([I-D.ietf-6tisch-terminology]). As opposed to a simple circuit that is a sequence of nodes and links, a Track is shaped as a directed acyclic graph towards a destination to support multi-path forwarding and route around failures. A Track may also branch off and rejoin, for the purpose of the so-called Packet Replication and Elimination (PRE), over non congruent branches. PRE may be used to complement layer-2 Automatic Repeat reQuest (ARQ) to

meet industrial expectations in Packet Delivery Ratio (PDR), in particular when the Track extends beyond the 6TiSCH network.

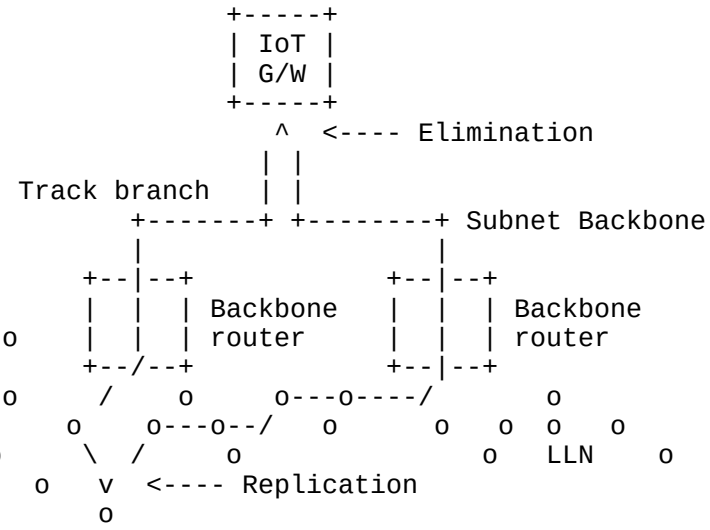


Figure 6: End-to-End deterministic Track

In the example above, 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.

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 makes it in due 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). It is also possible that the field device only uses the second branch if sending over the first branch fails.

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.

5.3.1. TSCH and 6top

6top 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 operations are specified in [I-D.wang-6tisch-6top-sublayer].

The 6top data model and management interfaces are further discussed in [I-D.ietf-6tisch-6top-interface] and [I-D.ietf-6tisch-coap].

The architecture defines "soft" cells and "hard" cells. "Hard" cells are owned and managed by an separate scheduling entity (e.g. a PCE) that specifies the slotOffset/channelOffset of the cells to be added/moved/deleted, in which case 6top can only act as instructed, and may not move hard cells in the TSCH schedule on its own.

5.3.2. SlotFrames and Priorities

A slotFrame is the base object that the PCE needs to manipulate to program a schedule into an LLN node. Elaboration on that concept can be found in section "SlotFrames and Priorities" of the 6TiSCH architecture [I-D.ietf-6tisch-architecture]. The architecture also details how the schedule is constructed and how transmission resources called cells can be allocated to particular transmissions so as to avoid collisions.

5.3.3. Schedule Management by a PCE

6TiSCH supports a mixed model of centralized routes and distributed routes. Centralized routes can for example be computed by a entity such as a PCE. Distributed routes are computed by RPL.

Both methods may inject routes in the Routing Tables of the 6TiSCH routers. In either case, each route is associated with a 6TiSCH topology that can be a RPL Instance topology or a track. The 6TiSCH topology is indexed by a Instance ID, in a format that reuses the RPLInstanceID as defined in RPL [RFC6550].

Both RPL and PCE rely on shared sources such as policies to define Global and Local RPLInstanceIDs that can be used by either method. It is possible for centralized and distributed routing to share a same topology. Generally they will operate in different slotFrames, and centralized routes will be used for scheduled traffic and will have precedence over distributed routes in case of conflict between the slotFrames.

Section "Schedule Management Mechanisms" of the 6TiSCH architecture describes 4 paradigms to manage the TSCH schedule of the LLN nodes: Static Scheduling, neighbor-to-neighbor Scheduling, remote monitoring and scheduling management, and Hop-by-hop scheduling. The Track operation for DetNet corresponds to a remote monitoring and scheduling management by a PCE.

The 6top interface document [I-D.ietf-6tisch-6top-interface] specifies the generic data model that can be used to monitor and manage resources of the 6top sublayer. Abstract methods are suggested for use by a management entity in the device. The data model also enables remote control operations on the 6top sublayer.

[I-D.ietf-6tisch-coap] defines an mapping of the 6top set of commands, which is described in [I-D.ietf-6tisch-6top-interface], to CoAP resources. This allows an entity to interact with the 6top layer of a node that is multiple hops away in a RESTful fashion.

[I-D.ietf-6tisch-coap] also defines a basic set CoAP resources and associated RESTful access methods (GET/PUT/POST/DELETE). The payload (body) of the CoAP messages is encoded using the CBOR format. The PCE commands are expected to be issued directly as CoAP requests or to 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.

5.3.4. Track Forwarding

By forwarding, this specification means the per-packet operation that allows to deliver a packet to a next hop or an upper layer in this node. Forwarding is based on pre-existing state that was installed as a result of the routing computation of a Track by a PCE. The 6TiSCH architecture supports three different forwarding model, G-MPLS Track Forwarding (TF), 6LoWPAN Fragment Forwarding (FF) and IPv6 Forwarding (6F) which is the classical IP operation. The DetNet case relates to the Track Forwarding operation under the control of a PCE.

A Track is a unidirectional path between a source and a destination. In a Track cell, the normal operation of IEEE802.15.4 Automatic Repeat-reQuest (ARQ) usually happens, though the acknowledgment may be omitted in some cases, for instance if there is no scheduled cell for a retry.

Track Forwarding is the simplest and fastest. A bundle of cells set to receive (RX-cells) is uniquely paired to a bundle of cells that

are set to transmit (TX-cells), representing a layer-2 forwarding state that can be used regardless of the network layer protocol. This model can effectively be seen as a Generalized Multi-protocol Label Switching (G-MPLS) operation in that the information used to switch a frame is not an explicit label, but rather related to other properties of the way the packet was received, a particular cell in the case of 6TiSCH. As a result, as long as the TSCH MAC (and Layer-2 security) accepts a frame, that frame can be switched regardless of the protocol, whether this is an IPv6 packet, a 6LoWPAN fragment, or a frame from an alternate protocol such as WirelessHART or ISA100.11a.

A data frame that is forwarded along a Track normally has a destination MAC address that is set to broadcast - or a multicast address depending on MAC support. This way, the MAC layer in the intermediate nodes accepts the incoming frame and 6top switches it without incurring a change in the MAC header. In the case of IEEE802.15.4, this means effectively broadcast, so that along the Track the short address for the destination of the frame is set to 0xFFFF.

A Track is thus formed end-to-end as a succession of paired bundles, a receive bundle from the previous hop and a transmit bundle to the next hop along the Track, and a cell in such a bundle belongs to at most one Track. For a given iteration of the device schedule, the effective channel of the cell is obtained by adding a pseudo-random number to the channelOffset of the cell, which results in a rotation of the frequency that used for transmission. The bundles may be computed so as to accommodate both variable rates and retransmissions, so they might not be fully used at a given iteration of the schedule. The 6TiSCH architecture provides additional means to avoid waste of cells as well as overflows in the transmit bundle, as follows:

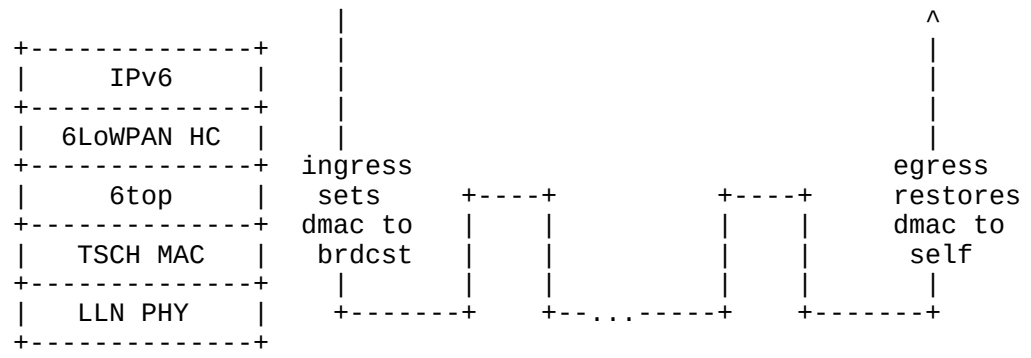
In one hand, a TX-cell that is not needed for the current iteration may be reused opportunistically on a per-hop basis for routed packets. When all of the frame that were received for a given Track are effectively transmitted, any available TX-cell for that Track can be reused for upper layer traffic for which the next-hop router matches the next hop along the Track. In that case, the cell that is being used is effectively a TX-cell from the Track, but the short address for the destination is that of the next-hop router. It results that a frame that is received in a RX-cell of a Track with a destination MAC address set to this node as opposed to broadcast must be extracted from the Track and delivered to the upper layer (a frame with an unrecognized MAC address is dropped at the lower MAC layer and thus is not received at the 6top sublayer).

On the other hand, it might happen that there are not enough TX-cells in the transmit bundle to accommodate the Track traffic, for instance if more retransmissions are needed than provisioned. In that case, the frame can be placed for transmission in the bundle that is used for layer-3 traffic towards the next hop along the track as long as it can be routed by the upper layer, that is, typically, if the frame transports an IPv6 packet. The MAC address should be set to the next-hop MAC address to avoid confusion. It results that a frame that is received over a layer-3 bundle may be in fact associated to a Track. In a classical IP link such as an Ethernet, off-track traffic is typically in excess over reservation to be routed along the non-reserved path based on its QoS setting. But with 6TiSCH, since the use of the layer-3 bundle may be due to transmission failures, it makes sense for the receiver to recognize a frame that should be re-tracked, and to place it back on the appropriate bundle if possible. A frame should be re-tracked if the Per-Hop-Behavior group indicated in the Differentiated Services Field in the IPv6 header is set to Deterministic Forwarding, as discussed in Section 5.4.1. A frame is re-tracked by scheduling it for transmission over the transmit bundle associated to the Track, with the destination MAC address set to broadcast.

There are 2 modes for a Track, transport mode and tunnel mode.

5.3.4.1. Transport Mode

In transport mode, the Protocol Data Unit (PDU) is associated with flow-dependant meta-data that refers uniquely to the Track, so the 6top sublayer can place the frame in the appropriate cell without ambiguity. In the case of IPv6 traffic, this flow identification is transported in the Flow Label of the IPv6 header. Associated with the source IPv6 address, the Flow Label forms a globally unique identifier for that particular Track that is validated at egress before restoring the destination MAC address (DMAC) and punting to the upper layer.



Track Forwarding, Transport Mode

5.3.4.2. Tunnel Mode

In tunnel mode, the frames originate from an arbitrary protocol over a compatible MAC that may or may not be synchronized with the 6TiSCH network. An example of this would be a router with a dual radio that is capable of receiving and sending WirelessHART or ISA100.11a frames with the second radio, by presenting itself as an access Point or a Backbone Router, respectively.

In that mode, some entity (e.g. PCE) can coordinate with a WirelessHART Network Manager or an ISA100.11a System Manager to specify the flows that are to be transported transparently over the Track.

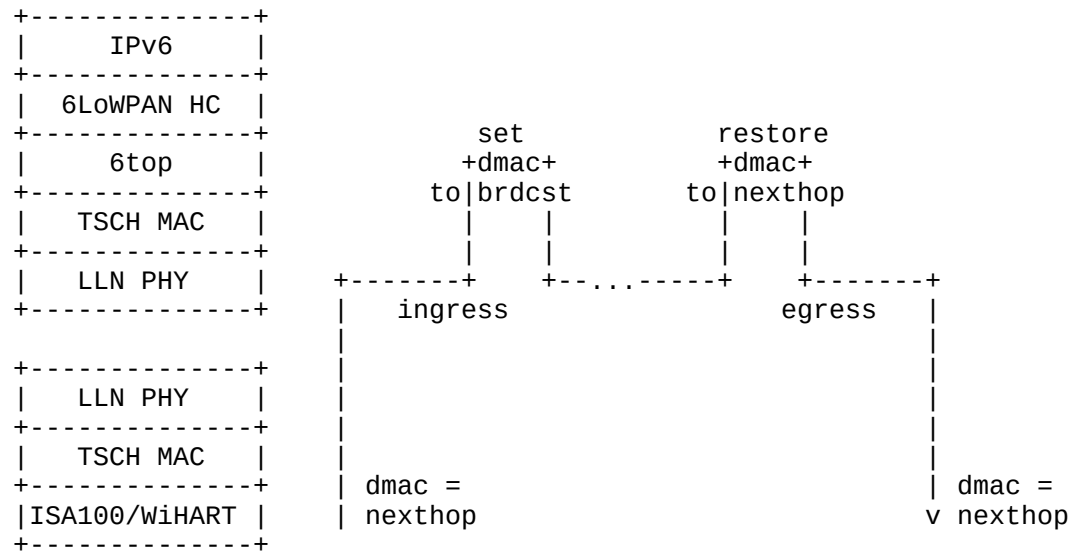


Figure 7: Track Forwarding, Tunnel Mode

In that case, the flow information that identifies the Track at the ingress 6TiSCH router is derived from the RX-cell. The dmac is set to this node but the flow information indicates that the frame must be tunneled over a particular Track so the frame is not passed to the upper layer. Instead, the dmac is forced to broadcast and the frame is passed to the 6top sublayer for switching.

At the egress 6TiSCH router, the reverse operation occurs. Based on metadata associated to the Track, the frame is passed to the appropriate link layer with the destination MAC restored.

5.3.4.3. Tunnel Metadata

Metadata coming with the Track configuration is expected to provide the destination MAC address of the egress endpoint as well as the tunnel mode and specific data depending on the mode, for instance a service access point for frame delivery at egress. If the tunnel egress point does not have a MAC address that matches the configuration, the Track installation fails.

In transport mode, if the final layer-3 destination is the tunnel termination, then it is possible that the IPv6 address of the destination is compressed at the 6LoWPAN sublayer based on the MAC address. It is thus mandatory at the ingress point to validate that the MAC address that was used at the 6LoWPAN sublayer for compression matches that of the tunnel egress point. For that reason, the node

that injects a packet on a Track checks that the destination is effectively that of the tunnel egress point before it overwrites it to broadcast. The 6top sublayer at the tunnel egress point reverts that operation to the MAC address obtained from the tunnel metadata.

5.4. Operations of Interest for DetNet and PCE

In a classical system, the 6TiSCH device does not place the request for bandwidth between self and another device in the network. Rather, an Operation Control System invoked through an Human/Machine Interface (HMI) indicates the Traffic Specification, in particular in terms of latency and reliability, and the end nodes. With this, the PCE must compute a Track 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 to recognize when a certain packet belongs to a certain Track, sort out duplicates, 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 Tracks. 6TiSCH expects that the programing of the schedule will be done over COAP as discussed in 6TiSCH Resource Management and Interaction using CoAP [I-D.ietf-6tisch-coap].

But an Hybrid mode may be required as well whereby a single Track is added, modified, or removed, for instance if it appears that a Track does not perform as expected for, say, PDR. For that case, the expectation is that a protocol that flows along a Track (to be), in a fashion similar to classical Traffic Engineering (TE) [CCAMP], may be used to update the state in the devices. 6TiSCH provides means for a device to negotiate a timeSlot with a neighbor, but in general that flow was not designed and no protocol was selected and it is expected that DetNet will determine the appropriate end-to-end protocols to be used in that case.

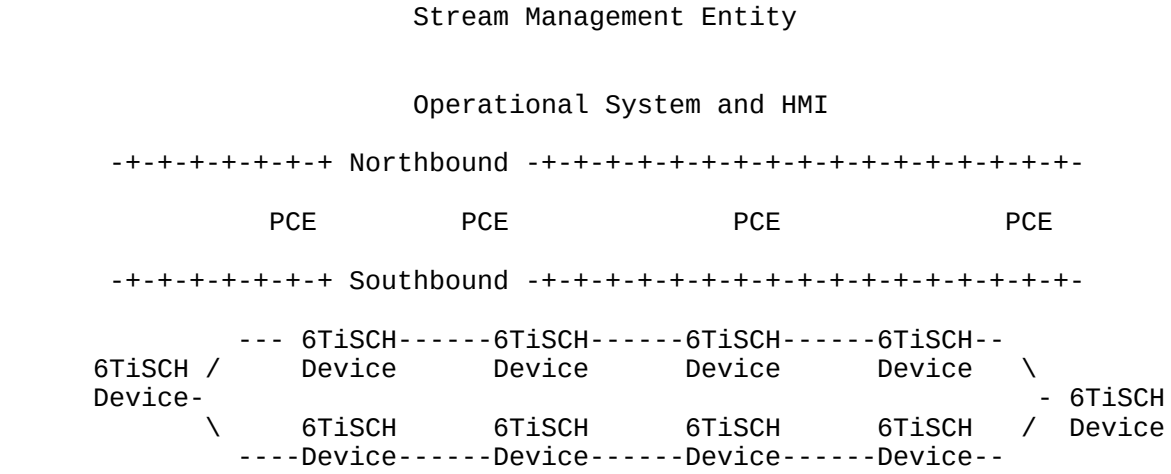


Figure 8

5.4.1. Packet Marking and Handling

Section "Packet Marking and Handling" of [I-D.ietf-6tisch-architecture] describes the packet tagging and marking that is expected in 6TiSCH networks.

5.4.1.1. Tagging Packets for Flow Identification

For packets that are routed by a PCE along a Track, the tuple formed by the IPv6 source address and a local RPLInstanceID is tagged in the packets to identify uniquely the Track and associated transmit bundle of timeSlots.

It results that the tagging that is used for a DetNet flow outside the 6TiSCH LLN MUST be swapped into 6TiSCH formats and back as the packet enters and then leaves the 6TiSCH network.

Note: The method and format used for encoding the RPLInstanceID at 6lo is generalized to all 6TiSCH topological Instances, which includes Tracks.

5.4.1.2. Replication, Retries and Elimination

6TiSCH expects elimination and replication of packets along a complex Track, but has no position about how the sequence numbers would be tagged in the packet.

As it goes, 6TiSCH expects that timeSlots corresponding to copies of a same packet along a Track are correlated by configuration, and does not need to process the sequence numbers.

The semantics of the configuration MUST enable correlated timeSlots to be grouped for transmit (and respectively receive) with a 'OR' relations, and then a 'AND' relation MUST be configurable between groups. The semantics is that if the transmit (and respectively receive) operation succeeded in one timeSlot in a 'OR' group, then all the other timeSLots in the group are ignored. Now, if there are at least two groups, the 'AND' relation between the groups indicates that one operation must succeed in each of the groups.

On the transmit side, timeSlots provisioned for retries along a same branch of a Track are placed a same 'OR' group. The 'OR' relation indicates that if a transmission is acknowledged, then further transmissions SHOULD NOT be attempted for timeSlots in that group. There are as many 'OR' groups as there are branches of the Track departing from this node. Different 'OR' groups are programmed for the purpose of replication, each group corresponding to one branch of the Track. The 'AND' relation between the groups indicates that transmission over any of branches MUST be attempted regardless of whether a transmission succeeded in another branch. It is also possible to place cells to different next-hop routers in a same 'OR' group. This allows to route along multi-path tracks, trying one next-hop and then another only if sending to the first fails.

On the receive side, all timeSlots are programmed in a same 'OR' group. Retries of a same copy as well as converging branches for elimination are converged, meaning that the first successful reception is enough and that all the other timeSlots can be ignored.

5.4.1.3. Differentiated Services Per-Hop-Behavior

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

5.4.2. Topology and capabilities

6TiSCH nodes are usually IoT devices, characterized by very limited amount of memory, just enough buffers to store one or a few IPv6 packets, and limited bandwidth between peers. It results that a node will maintain only a small number of peering information, and will not be able to store many packets waiting to be forwarded. Peers can be identified through MAC or IPv6 addresses, but a Cryptographically Generated Address [RFC3972] (CGA) may also be used.

Neighbors can be discovered over the radio using mechanism such as beacons, but, though the neighbor information is available in the 6TiSCH interface data model, 6TiSCH does not describe a protocol to pro-actively push the neighborhood information to a PCE. This protocol should be described and should operate over CoAP. The protocol should be able to carry multiple metrics, in particular the same metrics as used for RPL operations [RFC6551]

The energy that the device consumes in sleep, transmit and receive modes can be evaluated and reported. So can the amount of energy that is stored in the device and the power that it can be scavenged from the environment. The PCE SHOULD be able to compute Tracks that will implement policies on how the energy is consumed, for instance balance between nodes, ensure that the spent energy does not exceeded the scavenged energy over a period of time, etc...

5.5. Security Considerations

On top of the classical protection of control signaling that can be expected to support DetNet, it must be noted that 6TiSCH networks operate on limited resources that can be depleted rapidly if an attacker manages to operate a DoS attack on the system, for instance by placing a rogue device in the network, or by obtaining management control and to setup extra paths.

5.6. Acknowledgments

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.

6. Cellular Radio Use Cases

(This section was derived from draft-korhonen-detnet-telreq-00)

6.1. Introduction and background

The recent developments in telecommunication networks, especially in the cellular domain, are heading towards transport networks where precise time synchronization support has to be one of the basic building blocks. While the transport networks themselves have practically transitioned to all-AP packet based networks to meet the bandwidth and cost requirements, a highly accurate clock distribution has become a challenge. Earlier the transport networks in the cellular domain were typically time division and 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) [SyncE] were used. 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. The time and synchronization support are already topical for backhaul and midhaul packet networks [MEF], and becoming a real issue for fronthaul networks. Specifically in the fronthaul networks the timing and synchronization requirements can be extreme for packet based technologies, for example, in order of sub +-20 ns packet delay variation (PDV) and frequency accuracy of +0.002 PPM [Fronthaul].

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 (RAN) [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][I-D.mirsky-mpls-residence-time], these solution are not necessarily sufficient for the forthcoming RAN architectures or guarantee the higher time-synchronization requirements [CPRI]. There are also existing solutions for the TDM over IP [RFC5087] [RFC4553] or Ethernet transports [RFC5086]. The really interesting and important existing 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. While IEEE 802.1AS [IEEE8021AS] specifies a Layer-2 time synchronizing service other specification, 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]. Similarly to IEEE 1722 there is an ongoing standardization effort to define Layer-2 transport encapsulation format for transporting radio over Ethernet (RoE) in IEEE 1904.3 Task Force [IEEE19043].

As already mentioned all-IP RANs and various "haul" 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, especially in MPLS/PWE space, to establish circuits through the routed and switched networks, there is a lack of signaling the time synchronization and time-sensitive stream requirements/reservations for Layer-3 flows in a way that the entire transport stack is addressed and the Ethernet layers that needs to be configured are addressed. Furthermore, not all "user plane" traffic will be IP. Therefore, the same solution need also address the use cases where the user plane traffic is again another layer or 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 eventually targets to provide solutions for time-sensitive (IP/ transport) streams with deterministic properties over Ethernet-based switched networks.

This document describes requirements for deterministic networking in a cellular telecom transport networks context. The requirements include time synchronization, clock distribution and ways of establishing time-sensitive streams for both Layer-2 and Layer-3 user plane traffic using IETF protocol solutions.

The recent developments in telecommunication networks, especially in the cellular domain, are heading towards transport networks where precise time synchronization support has to be one of the basic building blocks. While the transport networks themselves have practically transitioned to all-AP packet based networks to meet the bandwidth and cost requirements, a highly accurate clock distribution has become a challenge. Earlier the transport networks in the cellular domain were typically time division and 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) [SyncE] were used. 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. The time and synchronization support are already topical for backhaul and midhaul packet networks [MEF], and becoming a real issue for fronthaul

networks. Specifically in the fronthaul networks the timing and synchronization requirements can be extreme for packet based technologies, for example, in order of sub +-20 ns packet delay variation (PDV) and frequency accuracy of +0.002 PPM [Fronthaul].

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 (RAN) [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][I-D.mirsky-mpls-residence-time], these solution are not necessarily sufficient for the forthcoming RAN architectures or guarantee the higher time-synchronization requirements [CPRI]. There are also existing solutions for the TDM over IP [RFC5087] [RFC4553] or Ethernet transports [RFC5086]. The really interesting and important existing 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. While IEEE 802.1AS [IEEE8021AS] specifies a Layer-2 time synchronizing service other specification, 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]. Similarly to IEEE 1722 there is an ongoing standardization effort to define Layer-2 transport encapsulation format for transporting radio over Ethernet (RoE) in IEEE 1904.3 Task Force [IEEE19043].

As already mentioned all-IP RANs and various "haul" 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, especially in MPLS/PWE space, to establish circuits through the routed and switched networks, there is a lack of signaling the time synchronization and time-sensitive stream requirements/reservations for Layer-3 flows in a way that the entire transport stack is addressed and the Ethernet layers that needs to be configured are addressed. Furthermore, not all "user plane" traffic will be IP. Therefore, the same solution need also address the use cases where the user plane traffic is again another layer or 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 eventually targets to provide solutions for time-sensitive (IP/ transport) streams with deterministic properties over Ethernet-based switched networks.

This document describes requirements for deterministic networking in a cellular telecom transport networks context. The requirements include time synchronization, clock distribution and ways of establishing time-sensitive streams for both Layer-2 and Layer-3 user plane traffic using IETF protocol solutions.

6.2. Network architecture

Figure Figure 9 illustrates a typical, 3GPP defined, cellular network architecture, which also has fronthaul and midhaul network segments. The fronthaul refers to the network connecting base stations (base band processing units) to the remote radio heads (antennas). The midhaul network typically refers to the network inter-connecting base stations (or small/pico cells).

Fronthaul networks build on the available excess time after the base band processing of the radio frame has completed. Therefore, the available time for networking is actually very limited, which in practise determines how far the remote radio heads can be from the base band processing units (i.e. base stations). For example, in a case of LTE radio the Hybrid ARQ processing of a radio frame is allocated 3 ms. Typically the processing completes way earlier (say up to 400 us, could be much less, though) thus allowing the remaining time to be used e.g. for fronthaul network. 200 us equals roughly 40 km of optical fiber based transport (assuming round trip time would be total 2*200 us). The base band processing time and the available "delay budget" for the fronthaul is a subject to change, possibly dramatically, in the forthcoming "5G" to meet, for example, the envisioned reduced radio round trip times, and other architecural and service requirements [NGMN].

The maximum "delay budget" is then consumed by all nodes and required buffering between the remote radio head and the base band processing in addition to the distance incurred delay. Packet delay variation (PDV) is problematic to fronthaul networks and must be minimized. If the transport network cannot guarantee low enough PDV additional buffering has to be introduced at the edges of the network to buffer out the jitter. Any buffering will eat up the total available delay budget, though. Section Section 6.3 will discuss the PDV requirements in more detail.

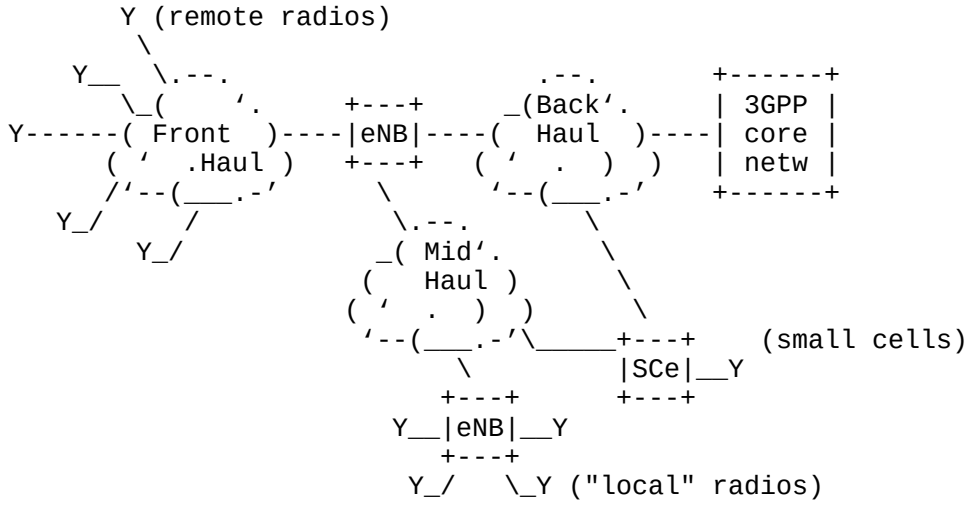


Figure 9: Generic 3GPP-based cellular network architecture with Front/Mid/Backhaul networks

6.3. Time synchronization requirements

Cellular networks starting from long term evolution (LTE) [TS36300] [TS23401] radio the phase synchronization is also needed in addition to the frequency synchronization. The commonly referenced fronthaul network synchronization requirements are typically drawn from the common public radio interface (CPRI) [CPRI] specification that defines the transport protocol between the base band processing - radio equipment controller (REC) and the remote antenna - radio equipment (RE). However, the fundamental requirements still originate from the respective cellular system and radio specifications such as the 3GPP ones [TS25104][TS36104][TS36211] [TS36133].

The fronthaul time synchronization requirements for the current 3GPP LTE-based networks are listed below:

Transport link contribution to radio frequency error:

+-2 PPB. The given value is considered to be "available" for the fronthaul link out of the total 50 PPB budget reserved for the radio interface.

Delay accuracy:

+-8.138 ns i.e. +-1/32 Tc (UMTS Chip time, Tc, 1/3.84 MHz) to downlink direction and excluding the (optical) cable length in one

direction. Round trip accuracy is then +-16.276 ns. The value is this low to meet the 3GPP timing alignment error (TAE) measurement requirements.

Packet delay variation (PDV):

* 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.

The above listed time synchronization requirements are hard to meet even with point to point connected networks, not to mention cases where the underlying transport network actually constitutes of multiple hops. It is expected that network deployments have to deal with the jitter requirements buffering at the very ends of the connections, 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 are valuable to make it easier to meet the end to end requirements.

In order to meet the timing requirements both senders and receivers must is perfect sync. This asks for a very accurate clock distribution solution. Basically all means and hardware support for guaranteeing accurate time synchronization in the network is needed. As an example support for 1588 transparent clocks (TC) in every intermediate node would be helpful.

6.4. Time-sensitive stream requirements

In addition to the time synchronization requirements listed in Section Section 6.3 the fronthaul networks assume practically error free transport. The maximum bit error rate (BER) has been defined to be 10^-12. When packetized that would equal roughly to packet error rate (PER) of 2.4*10^-9 (assuming ˜300 bytes packets). Retransmitting lost packets and/or using forward error coding (FEC) to circumvent bit errors are practically impossible due additional incurred delay. Using redundant streams for better guarantees for

delivery is also practically impossible due to high bandwidth requirements fronthaul networks have. For instance, current uncompressed CPRI bandwidth expansion ratio is roughly 20:1 compared to the IP layer user payload it carries in a "radio sample form".

The other fundamental assumption is that fronthaul links are symmetric. Last, all fronthaul streams (carrying radio data) have equal priority and cannot delay or pre-empt each other. This implies the network has always be sufficiently under subscribed to guarantee each time-sensitive flow meets their schedule.

Mapping the fronthaul requirements to [I-D.finn-detnet-architecture] Section 3 "Providing the DetNet Quality of Service" what is seemed usable are:

(a) Zero congestion loss.

(b) Pinned-down paths.

The current time-sensitive networking features may still not be sufficient for fronthaul traffic. Therefore, having specific profiles that take the requirements of fronthaul into account are deemed to be useful [IEEE8021CM].

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 not limited to solutions directly over Ethernet, over IP, and MPLS/PseudoWire transport.

6.5. Security considerations

Establishing time-sensitive streams in the network entails reserving networking resources sometimes for a considerable long time. It is important that these reservation requests must be authenticated to prevent malicious reservation attempts from hostile nodes or even accidental misconfiguration. This is specifically important in a case where the reservation requests span administrative domains. Furthermore, the reservation information itself should be digitally signed to reduce the risk where a legitimate node pushed a stale or hostile configuration into the networking node.

7. Other Use Cases

(This section was derived from draft-zha-detnet-use-case-00)

7.1. Introduction

The rapid growth of the today’s communication system and its access into almost all aspects of daily life has led to great dependency on services it provides. The communication network, as it is today, has applications such as multimedia and peer-to-peer file sharing distribution that require Quality of Service (QoS) guarantees in terms of delay and jitter to maintain a certain level of performance. Meanwhile, mobile wireless communications has become an important part to support modern sociality with increasing importance over the last years. A communication network of hard real-time and high reliability is essential for the next concurrent and next generation mobile wireless networks as well as its bearer network for E-2-E performance requirements.

Conventional transport network is IP-based because of the bandwidth and cost requirements. However the delay and jitter guarantee becomes a challenge in case of contention since the service here is not deterministic but best effort. With more and more rigid demand in latency control in the future network [METIS], deterministic networking [I-D.finn-detnet-architecture] is a promising solution to meet the ultra low delay applications and use cases. There are already typical issues for delay sensitive networking requirements in midhaul and backhaul network to support LTE and future 5G network [net5G]. And not only in the telecom industry but also other vertical industry has increasing demand on delay sensitive communications as the automation becomes critical recently.

More specifically, CoMP techniques, D-2-D, industrial automation and gaming/media service all have great dependency on the low delay communications as well as high reliability to guarantee the service performance. Note that the deterministic networking is not equal to low latency as it is more focused on the worst case delay bound of the duration of certain application or service. It can be argued that without high certainty and absolute delay guarantee, low delay provisioning is just relative [rfc3393], which is not sufficient to some delay critical service since delay violation in an instance cannot be tolerated. Overall, the requirements from vertical industries seem to be well aligned with the expected low latency and high determinist performance of future networks

This document describes several use cases and scenarios with requirements on deterministic delay guarantee within the scope of the deterministic network [I-D.finn-detnet-problem-statement].

7.2. Critical Delay Requirements

Delay and jitter requirement has been take into account as a major component in QoS provisioning since the birth of Internet. The delay sensitive networking with increasing importance become the root of mobile wireless communications as well as the applicable areas which are all greatly relied on low delay communications. Due to the best effort feature of the IP networking, mitigate contention and buffering is the main solution to serve the delay sensitive service. More bandwidth is assigned to keep the link low loaded or in another word, reduce the probability of congestion. However, not only lack of determinist but also has limitation to serve the applications in the future communication system, keeping low loaded cannot provide deterministic delay guarantee. Take the [METIS] that documents the fundamental challenges as well as overall technical goal of the 5G mobile and wireless system as the starting point. It should supports: -1000 times higher mobile data volume per area, -10 times to 100 times higher typical user data rate, -10 times to 100 times higher number of connected devices, -10 times longer battery life for low power devices, and -5 times reduced End-to-End (E2E) latency, at similar cost and energy consumption levels as today’s system. Taking part of these requirements related to latency, current LTE networking system has E2E latency less than 20ms [LTE-Latency] which leads to around 5ms E2E latency for 5G networks. It has been argued that fulfill such rigid latency demand with similar cost will be most challenging as the system also requires 100 times bandwidth as well as 100 times of connected devices. As a result to that, simply adding redundant bandwidth provisioning can be no longer an efficient solution due to the high bandwidth requirements more than ever before. In addition to the bandwidth provisioning, the critical flow within its reserved resource should not be affected by other flows no matter the pressure of the network. Robust defense of critical flow is also not depended on redundant bandwidth allocation. Deterministic networking techniques in both layer-2 and layer-3 using IETF protocol solutions can be promising to serve these scenarios.

7.3. Coordinated multipoint processing (CoMP)

In the wireless communication system, Coordinated multipoint processing (CoMP) is considered as an effective technique to solve the inter-cell interference problem to improve the cell-edge user throughput [CoMP].

7.3.1. CoMP Architecture

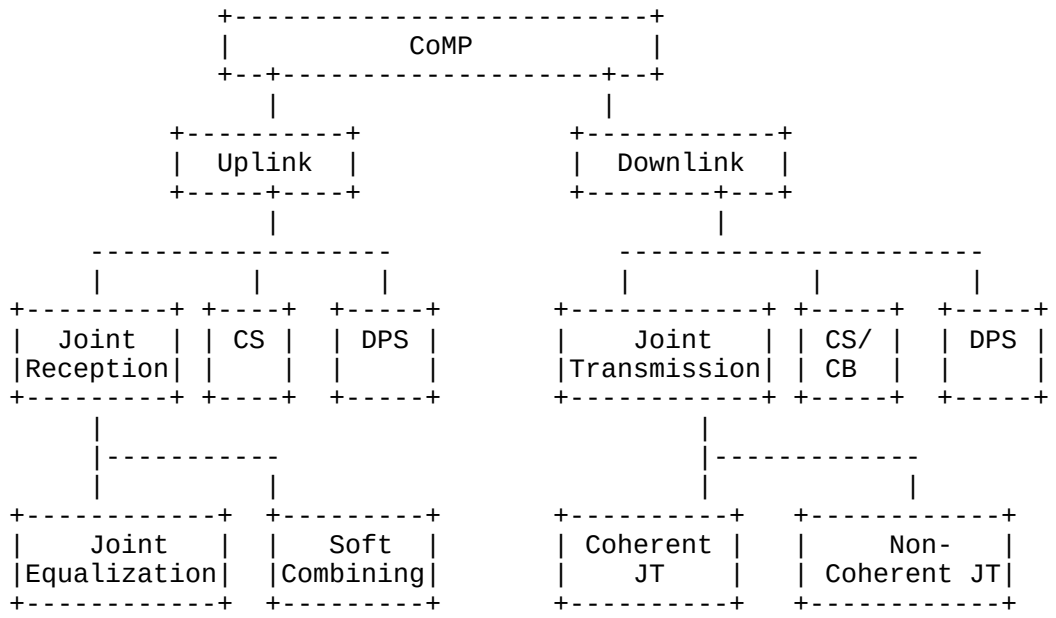


Figure 10: Framework of CoMP Technology

As shown in Figure 10, CoMP reception and transmission is a framework that 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-UEs transmission to a multi-cell- to-multi-UEs transmission by base station cooperation. In contrast to single-cell scenario, CoMP has critical issues such as: Backhaul latency, CSI (Channel State Information) reporting and accuracy and Network complexity. Clearly the first two requirements are very much delay sensitive and will be discussed in next section.

7.3.2. Delay Sensitivity in CoMP

As the essential feature of CoMP, signaling is exchanged between eNBs, the backhaul latency is the dominating limitation of the CoMP performance. Generally, JT and JP may benefit from coordinating the scheduling (distributed or centralized) of different cells in case that the signaling exchanging between eNBs is limited to 4-10ms. For C-RAN the backhaul latency requirement is 250us while for D-RAN it is 4-15ms. And this delay requirement is not only rigid but also absolute since any uncertainty in delay will down the performance significantly. Note that, some operator’s transport network is not build to support Layer-3 transfer in aggregation layer. In such case, the signaling is exchanged through EPC which means delay is

supposed to be larger. CoMP has high requirement on delay and reliability which is lack by current mobile network systems and may impact the architecture of the mobile network.

7.4. Industrial Automation

Traditional "industrial automation" terminology usually refers to automation of manufacturing, quality control and material processing. "Industrial internet" and "industrial 4.0" [EA12] is becoming a hot topic based on the Internet of Things. This high flexible and dynamic engineering and manufacturing will result in a lot of so-called smart approaches such as Smart Factory, Smart Products, Smart Mobility, and Smart Home/Buildings. No doubt that ultra high reliability and robustness is a must in data transmission, especially in the closed loop automation control application where delay requirement is below 1ms and packet loss less than 10E-9. All these critical requirements on both latency and loss cannot be fulfilled by current 4G communication networks. Moreover, the collaboration of the industrial automation from remote campus with cellular and fixed network has to be built on an integrated, cloud-based platform. In this way, the deterministic flows should be guaranteed regardless of the amount of other flows in the network. The lack of this mechanism becomes the main obstacle in deployment on of industrial automation.

7.5. Vehicle to Vehicle

V2V communication has gained more and more attention in the last few years and will be increasingly growth in the future. Not only equipped with direct communication system which is short ranged, V2V communication also requires wireless cellular networks to cover wide range and more sophisticated services. V2V application in the area autonomous driving has very stringent requirements of latency and reliability. It is critical that the timely arrival of information for safety issues. In addition, due to the limitation of processing of individual vehicle, passing information to the cloud can provide more functions such as video processing, audio recognition or navigation systems. All of those requirements lead to a highly reliable connectivity to the cloud. On the other hand, it is natural that the provisioning of low latency communication is one of the main challenges to be overcome as a result of the high mobility, the high penetration losses caused by the vehicle itself. As result of that, the data transmission with latency below 5ms and a high reliability of PER below 10E-6 are demanded. It can benefit from the deployment of deterministic networking with high reliability.

7.6. Gaming, Media and Virtual Reality

Online gaming and cloud gaming is dominating the gaming market since it allow multiple players to play together with more challenging and competing. Connected via current internet, the latency can be a big issue to degrade the end users' experience. There different types of games and FPS (First Person Shooting) gaming has been considered to be the most latency sensitive online gaming due to the high requirements of timing precision and computing of moving target. Virtual reality is also receiving more interests than ever before as a novel gaming experience. The delay here can be very critical to the interacting in the virtual world. Disagreement between what is seeing and what is feeling can cause motion sickness and affect what happens in the game. Supporting fast, real-time and reliable communications in both PHY/MAC layer, network layer and application layer is main bottleneck for such use case. The media content delivery has been and will become even more important use of Internet. Not only high bandwidth demand but also critical delay and jitter requirements have to be taken into account to meet the user demand. To make the smoothness of the video and audio, delay and jitter has to be guaranteed to avoid possible interruption which is the killer of all online media on demand service. Now with 4K and 8K video in the near future, the delay guarantee become one of the most challenging issue than ever before. 4K/8K UHD video service requires 6Gbps-100Gbps for uncompressed video and compressed video starting from 60Mbps. The delay requirement is 100ms while some specific interactive applications may require 10ms delay [UHD-video].

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. Acknowledgments

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.

10. 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>.

[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-02 (work in progress), November 2015.

[I-D.finn-detnet-problem-statement]
Finn, N. and P. Thubert, "Deterministic Networking Problem Statement", draft-finn-detnet-problem-statement-04 (work in progress), October 2015.

[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-08 (work in progress), May 2015.

[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-06 (work in progress), November 2015.

[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-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.mirsky-mpls-residence-time]
Mirsky, G., Ruffini, S., Gray, E., Drake, J., Bryant, S., and S. Vainshtein, "Residence Time Measurement in MPLS network", draft-mirsky-mpls-residence-time-07 (work in progress), July 2015.

[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>.

[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>.

[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