Deterministic Networks Gap Analysis
draft-dujovne-detnet-gap-analysis-00

Abstract

   This document introduces and describes several conditions and use
   cases where the use of an IP-based layer-3 and up is required to
   provide a complete networking solution to deterministic networks.
   The contents of this work is a gap analysis to contribute to the
   design and development of a number of complimentary modules to
   provide IP-enabled networking for deterministic networks.

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 10, 2015.

Copyright Notice

Table of Contents

1. Introduction

   The first step to become detnet networks was taken at the stub
   networks such as 6TiSCH-based ones [I-D.ietf-6tisch-tsch], meaning
   those simple networks that connect the endpoints to one or several
   gateways, thus enabling both a predictable delay and a very high
   reliability. These two characteristics must be preserved along the
   the multi-hop path between source and destination, thus requiring a
   fully-fledged deterministic end-to-end network including management
   services to achieve this goal. This is achieved by installing tracks
   with all the deterministic capabilities.

   The evolution towards these kinds of intermediate and backbone
   deterministic networks has been taken up to the MAC layer by the IEEE
   by defining several standards to provide building blocks to guarantee
   a predictable delay, such as buffers, queues and schedulers. The
   configuration, use, coordination, resources management and control of
   these blocks must be achieved at a higher layer, tightly linked to
   the routing scheme. Many applications are currently lacking this
   kind of solution, forcing bandwidth overprovisioning to reduce packet
   loss and delay uncertainty.

2. Assumptions

   Current packet loss and delay jitter provided by IP networks are not
   enough for industrial applications

   Realtime audio and video for reliable content distribution in local
   networks is cannot be achieved without time-scheduled IP networks

   End-to-end delay and packet loss guarantees cannot be provided
   without a managed deterministic network

3. Gap Analysis

3.1. 6TiSCH Track management

   6TiSCH requires the installation of tracks along a path with
   deterministic capabilities, including scheduled transmissions,
   intermediate queue management, synchronization and path and packet
   redundnacy among others. The mechanism to achieve such a path is
   achieved by using PCE/SDN operations as defined on RFC 7149
   [RFC7149].

   On 6TiSCH, IPv6 packets are carried on installed tracks; to reduce
   resource usage, there is also the need to forward IPv6 packets by
   opportunistic reuse of track slots and also reuse link bundle slots
   to forward schedule packets that missed their track. Both mechanisms
   require the use of Deterministic Networking managament capabilities.

3.2. Deterministic Payload on MPLS

   There are certain Non-IP protocols such as Profibus and Modbus which
   can be carried as IPv6 payload as long as this traffic is treated as
   Deterministic; this can be achieved by the use of MPLS and and a
   specific mangement module for layer-2 path redundancy, such as
   Parallel Redundancy Protocol.

3.3. Traffic Specification implementation for PCE

   There are several issues on the implementation of Traffic
   Specification for the PCE
   [I-D.ietf-teas-interconnected-te-info-exchange]:

      A TEAS adaptation to carry the topology (neighbors, link quality,
      interference test, etc.) and capabilities (buffers, queues and
      timers) from the point of view of the individual devices.

      A CCAMP/RSVP-TE adaptation to program the individual tracks

   An adaptation of PCEP to push an individual device schedule

3.4. Packet-track ID

   Track ID on packets is not defined yet; the use of Diffserv/DSCP and
   MPLS (and G-MPLS for 6TiSCH) are possible alternatives

3.5. Packet Redundancy Protocol

   There is a need to define a Packet Redundancy Protocol (PRP) for
   deterministic networks, including the PRP sequence number which can
   be defined by the ASN.

4. Specific requirements

4.1. Industrial Control

   TODO

4.2. Audio/Video distribution

   TODO

4.3. Resource Management

   TODO

4.4. Instrumentation (Technical/Scientific) networks

   TODO

4.5. Vehicular control

   TODO

5. Acknowledgments

   Thanks to the Fondecyt (CONICYT-Chile) 11121475 Project.

6. Informative References

   [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-05 (work in
              progress), January 2015.

   [I-D.ietf-pce-architecture]
            Farrel, A., "A Path Computation Element (PCE) Based
            Architecture", draft-ietf-pce-architecture-05 (work in
            progress), April 2006.

   [I-D.ietf-teas-interconnected-te-info-exchange]
            Farrel, A., Drake, J., Bitar, N., Swallow, G., Ceccarelli,
            D., and X. Zhang, "Problem Statement and Architecture for
            Information Exchange Between Interconnected Traffic
            Engineered Networks", draft-ietf-teas-interconnected-te-
            info-exchange-02 (work in progress), March 2015.

   [RFC7149] Boucadair, M. and C. Jacquenet, "Software-Defined
            Networking: A Perspective from within a Service Provider
            Environment", RFC 7149, March 2014.

Author's Address

   Diego Dujovne (editor)
   Universidad Diego Portales
   Escuela de Informatica y Telecomunicaciones
   Av. Ejercito 441
   Santiago, Region Metropolitana
   Chile

   Phone: +56 (2) 676-8121
   Email: diego.dujovne@mail.udp.cl

DetNet   N. Finn
Internet-Draft   P. Thubert
Intended status: Standards Track   Cisco
Expires: September 10, 2015   M. Johas Teener
                                                           Broadcom
                                                       March 9, 2015

                    Deterministic Networking Architecture
                      draft-finn-detnet-architecture-01

Abstract

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

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   Operational Technology (OT) refers to industrial networks that are
   typically used for monitoring systems and supporting control loops,
   as well as movement detection systems for use in process control
   (i.e., process manufacturing) and factory automation (i.e., discrete
   manufacturing). Due to its different goals, OT has evolved in
   parallel but in a manner that is radically different from IT/ICT,
   focusing on highly secure, reliable and deterministic networks, with
   limited scalability over a bounded area.

   The convergence of IT and OT technologies, also called the Industrial
   Internet, represents a major evolution for both sides. The work has
   already started; in particular, the industrial automation space has
   been developing a number of Ethernet-based replacements for existing
   digital control systems, often not packet-based (fieldbus
   technologies).

   These replacements are meant to provide similar behavior as the
   incumbent protocols, and their common focus is to transport a fully
   characterized flow over a well-controlled environment (i.e., a
   factory floor), with a bounded latency, extraordinarily low frame
   loss, and a very narrow jitter. Examples of such protocols include
   PROFINET, ODVA Ethernet/IP, and EtherCAT.

   In parallel, the need for determinism in professional and home audio/
   video markets drove the formation of the Audio/Video Bridging (AVB)
   standards effort of IEEE 802.1. With the explosion of demand for
   connectivity and multimedia in transportation in general, the
   Ethernet AVB technology has become one of the hottest topics, in
   particular in the automotive connectivity. It is finding application
   in all elements of the vehicle from head units, to rear seat
   entertainment modules, to amplifiers and camera modules. While aimed
   at less critical applications than some industrial networks, AVB
   networks share the requirement for extremely low packet loss rates
   and ensured finite latency and jitter.

   Other instances of in-vehicle deterministic networks have arisen as
   well for control networks in cars, trains and buses, as well as
   avionics, with, for instance, the mission-critical "Avionics Full-
   Duplex Switched Ethernet" (AFDX) that was designed as part of the
   ARINC 664 standards. Existing automotive control networks such as
   the LIN, CAN and FlexRay standards were not designed to cover these
   increasing demands in terms of bandwidth and scalability that we see
   with various kinds of Driver Assistance Systems (DAS) and new
   multiplexing technologies based on Ethernet are now getting traction.

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) 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, Industrial Automation segregates the network along the broad lines of the Purdue Enterprise Reference Architecture (PERA), using different technologies at each level, and public infrastructures such as Electricity Automation require deterministic properties over the Wide Area. The realization is now coming that the convergence of IT and OT networks requires Layer-3, as well as Layer-2, capabilities.

The present architecture is the result of a collaboration of the IETF and the IEEE and implements an abstract model that can be applicable both at Layer-2 and Layer-3, and along segments of different technologie. With this new work, a path may span, for instance, across a (limited) number of 802.1 bridges and then a (limited) number of IP routers. In that example, the IEEE 802.1 bridges may be operating at Layer-2 over Ethernet whereas the IP routers may be 6TiSCH nodes operating at Layer-2 and/or Layer-3 over the IEEE 802.15.4e MAC.

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 4.7 also require time synchronization among relay systems. The means used to achieve time synchronization are not addressed in this document.

2. Terminology

The follwing special terms are used in this document in order to avoid the assumption that a given element in the archetecture does or does not have Internet Protocol stack, functions as a router or a bridge, or otherwise plays a particular role at Layer-3 or higher:

bridge
        A Customer Bridge as defined by IEEE 802.1Q
        [IEEE802.1Q-2011].

circuit
        A trail of configuration from talker to listener(s) through
        relay systems associated with a DetNet stream, required to
        deliver the benefits of DetNet.

end system

Commonly called a "host" in IETF documents, and an "end
station" is IEEE 802 documents. End systems of interest to
this document are talkers and listeners.

listener
        An end system capable of sinking a DetNet stream.

relay system
        A router or a bridge.

stream

        A DetNet stream is a sequence of packets from a single
        talker, through some number of relay systems to one or more
        listeners, that is limited by the talker in its maximum
        packet size and transmission rate, and can thus be ensured
        the DetNet Quality of Service (QoS) from the network.

talker
        An end system capable of sourcing a DetNet stream.

3. Providing the DetNet Quality of Service

   DetNet Quality of Service is expressed in terms of:

   o Minimum and maximum end-to-end latency from talker to listener;

   o Probability of loss of a packet, assuming the normal operation of
      the relay systems and links;

   o Probability of loss of a packet in the event of the failure of a
      relay system or link.

   It is a distinction of DetNet that it is concerned solely with worst-
   case values for all of the above parameters. 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.

   Three techniques are employed by DetNet to achieve these QoS
   parameters:

   a. Zero congestion loss (Section 3.1). Network resources such as
       link bandwidth, buffers, queues, shapers, and scheduled input/
       output slots are assigned in each relay system to the use of a
       specific DetNet stream or group of streams. Note that, given a
       finite amount of buffer space), zero congestion loss necessarily
       ensures a maximum end-to-end latency. Depending on the method
       employed, a minimum latency can also be achieved.

   b. Pinned-down paths (Section 3.2). Point-to-point paths or point-
       to-multipoint trees through the network from a talker to one or
       more listeners can be established, and DetNet streams assigned to
       follow a particular path or tree.

   c. Packet replication and deletion (Section 3.3). End systems and/
       or relay systems can sequence number, replicate, and eliminate
       replicated packets at multiple points in the network in order to
       ensure that one (or more) equipment failure events still leave at
       least one path intact for a DetNet stream.

   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:

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

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

   o 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. However, these methods generally work best in
   the absence of any significant amount of non-critical traffic in the
   network (if, indeed, such traffic is supported at all), or work only
   if the critical traffic constitutes only a small portion of the
   network's theoretical capacity, or work only if all systems are
   functioning properly, or in the absence of actions by end systems
   that disrupt the network's operations.

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

industry) to make selections among the available means of
implementing DetNet networks.

3.1. Zero Congestion Loss

The primary means by which DetNet achieves its QoS assurances is to
completely eliminate congestion at an output port as a cause of
packet loss. Given that a DetNet stream 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 talker, and
every relay system along the path to the listener (or nearly every
relay system -- see Section 4.5.2) be careful to regulate its output
to not exceed the data rate for any stream, except for brief perios
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
stream.

The low-level mechanisms described in Section 4.7 provide the
necessary regulation of transmissions by an edge system or relay
system to ensure zero congestion loss. Of course, the reservation of
the bandwidth and buffers for a stream requires the provisioning
described in Section 4.12.

3.2. Pinned-down paths

In networks controlled by typical peer-to-peer protocols such as IEEE
802.1 ISIS bridged networks or ETF 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. For this reason,
many real-time networks rely on physical rings of two-port devices,
with a relatively simple ring control protocol. This both minimizes
recovery time and easily supports redundant paths. Of course, this
comes at the cost 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 brief losses of connectivity, DetNet employs pinned-down
paths, where the path taken by a given DetNet stream does not change,
at least immediately, and likely not at all, in response to network
topology events. When combined with seamless redundancy

(Section 3.3), this results in a high likelihood of continuous
connectivity.

3.3. Seamless Redundancy

   After congestion loss has been eliminated, the most important causes
   of packet loss are random media and/or memory faults and equipment
   failures.

   Seamless redundancy involves three capabilities:

   o Adding sequence numbers to the packets of a DetNet stream.

   o Replicating these packets and, typically, sending them along at
      least two different paths to the listener(s).

   o Discarding duplicated packets.

   In the simplest case, this amounts to replicating each packet in a
   talker that has two interfaces, and conveying them through the
   network, along separate paths, to the similarly dual-homed listeners,
   that discard the extras. This ensures that one path (with zero
   congestion loss) remains, even if some relay system fails.

   Alternatively, relay systems 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 systems
   each replicate (R) the DetNet stream on input, sending the stream to
   both the other relay system and to the end system, and eliminated
   duplicates (E) on the output interface to the right-hand end system.
   Any one links in the network can fail, and the Detnet stream can
   still get through. Furthermore, two links can fail, as long as they
   are in different segments of the network.

```
            > > > > > > >   relay    > > > > > > >
           > /------------+ R system E +------------\ >
          > /              v + ^                \ >
end    R +              v | ^                + E end
system   +              v | ^                +   system
          > \              v + ^                / >
           > \------------+ R relay  E +------------/ >
            > > > > > > >   system   > > > > > > >
```
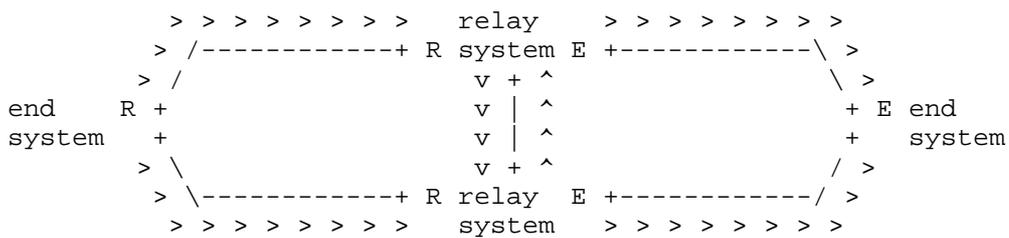
                          Figure 1

4. DetNet Architecture

   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 pinned-down 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 Software-Defined Networking (SDN): Layers
   and Architecture Terminology [RFC7426] which is represented below:

SDN Layers and Architecture Terminology per RFC 7426

```
               o------------------------------o
               |                              |
               | +-------------+  +----------+ |
               | | Application |  | Service  | |
               | +-------------+  +----------+ |
               |      Application Plane        |
               o---------------Y--------------o
                               |
  *----------------------------Y------------------------------*
  |        Network Services Abstraction Layer (NSAL)          |
  *------Y--------------------------------------------Y-------*
         |                                            |
         |              Service Interface             |
         |                                            |
  o------Y-----------------o      o-------------------Y------o
  |      |  Control Plane  |      | Management Plane  |      |
  | +----Y----+   +-----+  |      | +-----+   +----Y----+    |
  | | Service |   | App |  |      | | App |   | Service |    |
  | +----Y----+   +--Y--+  |      | +--Y--+   +----Y----+    |
  | |         |      |     |      | |      |        |        |
  | *----Y----------Y----* |      | *---Y-----------Y----*   |
  | | Control Abstraction | |      | | Management Abstraction | |
  | |    Layer (CAL)      | |      | |    Layer (MAL)      |   |
  | *----------Y---------* |      | *----------Y---------*   |
  | |          |          | |      | |          |          |  |
  o-----------|-----------o      o-----------|-------------o
              |                              |
              | CP                           | MP
              | Southbound                   | Southbound
              | Interface                    | Interface
              |                              |
  *-----------Y------------------------------Y----------------*
  |        Device and resource Abstraction Layer (DAL)        |
  *-----------Y------------------------------Y----------------*
  |           |              |               |                |
  |   o-------Y----------o  +-----+  o--------Y----------o     |
  |   | Forwarding Plane |  | App |  | Operational Plane |     |
  |   o-----------------o  +-----+  o-------------------o     |
  |                  Network Device                           |
  +-----------------------------------------------------------+
```

Figure 2

4.1. The Application Plane

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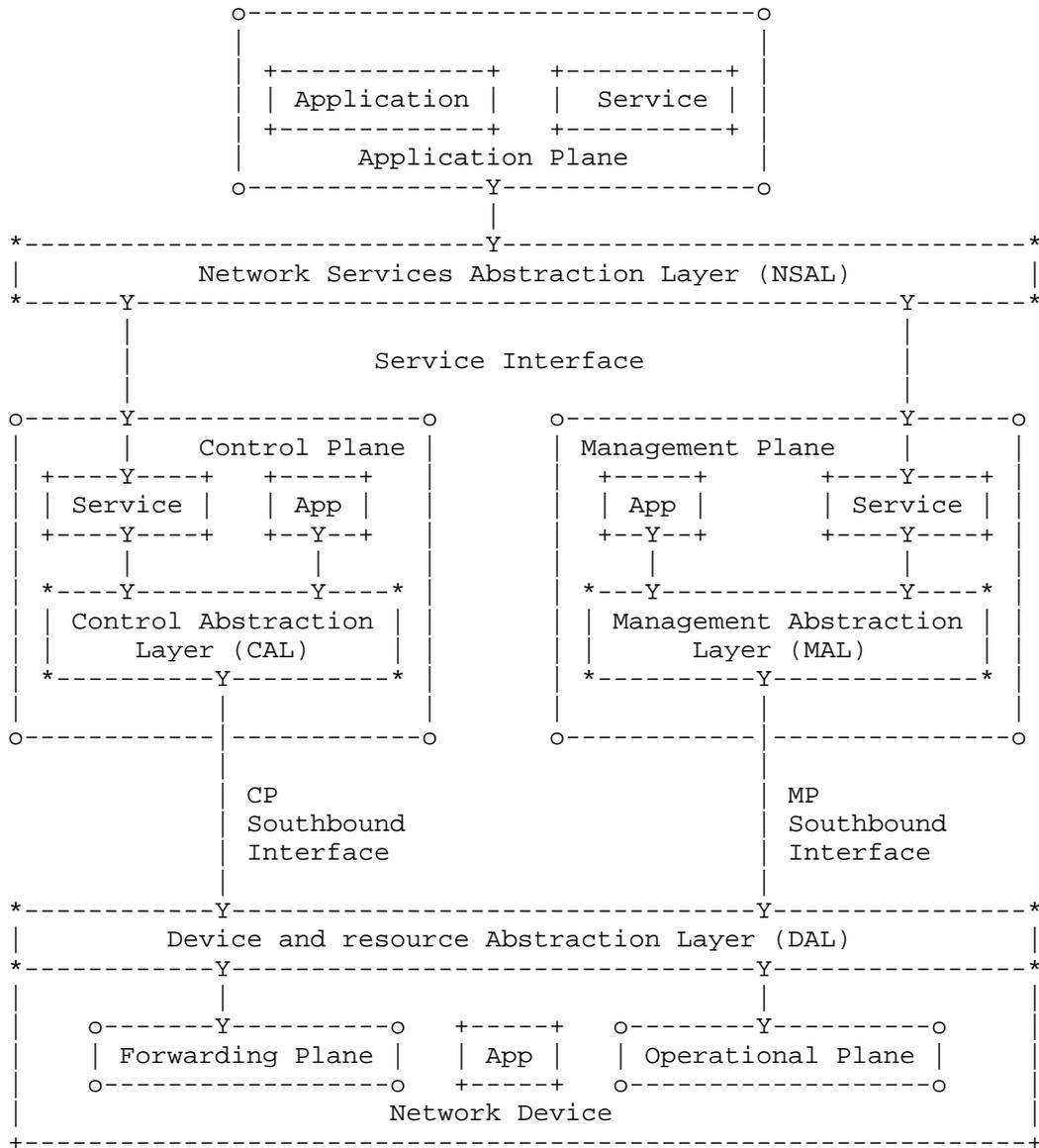   Agent, a specialized application that interacts with the end user /
   operator and performs requests for Deterministic Networking services
   via an abstract Stream Management Entity, (SME) which may or may not
   be collocated with (one of) the end systems.

   At the Application Plane, a management interface enables the
   negotiation of streams between end systems. An abstraction of the
   stream 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 relay systems.

4.2. 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 SME
   as Per-Stream Per-Hop Behaviors installed in the relay systems for
   each individual streams. The PCEs place each stream along a
   deterministic sequence of relay systems so as to respect per-stream
   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.

4.3. 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.

   The network Plane comprises the Network Interface Cards (NIC) in the
   end systems, which are typically IP hosts, and relay systems, which
   are typically IP routers and switches. Network-to-Network Interfaces
   such as used for Traffic Engineering path reservation in [RFC3209],
   as well as User-to-Network Interfaces (UNI) such as provided by the
   Local Management Interface (LMI) between network and end systems, are
   all part of the Network 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.

```
                        Stream Management Entity

      End                                                        End
      System                                                     System


    -+-+-+-+-+-+-+ Northbound -+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

            PCE           PCE                PCE             PCE


    -+-+-+-+-+-+-+ Southbound -+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

            Relay       Relay       Relay       Relay
            System      System      System      System
     NIC                                                        NIC
            Relay       Relay       Relay       Relay
            System      System      System      System
```
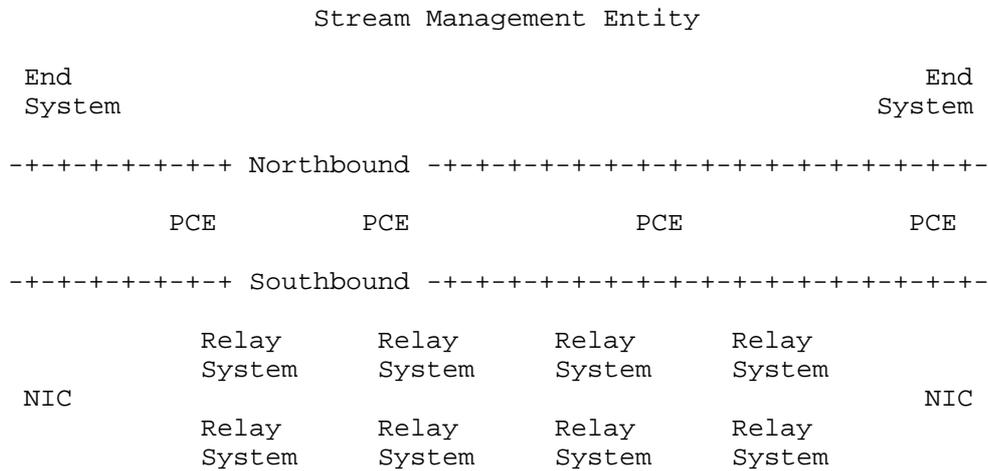
                              Figure 3

   The relay systems (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-stream
   paths up, providing a Stream Characterization that is more tightly
   coupled to the relay system Operation than a TSpec.

   At the Network plane, relay systems 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 stream, 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.

4.4. Elements of DetNet Architecture

The DetNet architecture has a number of elements, discussed in the
following sections:

a. A model for the definition, identification, and operation of
    DetNet streams (Section 4.5), for use by relay systems to
    classify and process individual packets following per-stream
    rules.

b. A model for the flow of data from an end system or through a
    relay system that can be used to predict the bounds for that
    system's impact on the QoS of a DetNet stream, without
    significantly constraining the method of implementing that
    system, for use by the Controllers to configure policing and
    shaping engines in Network Systems over the Southbound interface.
    The model includes:

    1. A model for queuing, transmission selection, shaping,
        preemption, and timing resources that can be used by an end
        system or relay system to control the selection of packets
        output on an interface. These models must have sufficiently
        well-defined characteristics, both individually and in the
        aggregate, to give predictable results for the QoS for DetNet
        packets (Section 4.7).

    2. A model for identifying misbehaving DetNet streams and
        mitigating their impact on properly functioning streams
        (Section 4.9).

c. A model for the relay system to inform the controller(s) of the
    information it needs for adequate path computations including:

    1. Systems' individual capabilities (e.g. can do replication,
        can do precise time).

    2. Link capabilities and resources (e.g. bandwidth, 0 delays,
        hardware deterministic support to the physical layer, ...)

    3. hysical resources (total and available buffers, timers,
        queues, etc)

4. Network Adjacencies (neighbors)

d. A model for the provision of a service, by end systems, or relay systems, to forward a DetNet stream over a simple or redundant path. The model includes:

1. A model for an abstract relaying operation of either Routing or forwarding packets of a DetNet stream to a next-hop relay system, across Layer boundaries.

2. A model of next-hop(s) information for replicating the packets of a DetNet stream, typically at or near the talker, merging and/or re-replicating those packets at other points in the network, and finally eliminating the duplicates, typically at or near the listener(s), in order to provide high availability (Section 3.3).

e. The protocol stack model for an end system and/or a relay system should support the above elements in a manner that maximizes the applicability of existing standards and protocols to the DetNet problem, allows for the creation of new protocols where needed, thus making DetNet an add-on feature to existing networks, rather than a new way to do networking. In particular this protocol stack supports networks in which the path from talker to listener(s) includes bridges and/or routers in any order (Section 4.10).

f. A variety of models for the provisioning of DetNet streams can be envisioned, including orchestration by a central controller or by a federation of controllers, provisioning by relay systems and end systems sharing peer-to-peer protocols, by off-line configuration, or by a combination of these methods. The provisioning models are similar to existing Layer-2 and Layer-3 models, in order to minimize the amount of innovation required in this area (Section 4.12).

4.5. DetNet streams

4.5.1. Talker guarantees

DetNet streams can by synchronous or asynchronous. The transmission of packets in synchronous DetNet streams uses time synchronization among the end and relay systems to control the flow of packets. Asynchronous DetNet streams are characterized by:

o A maximum packet size;

o An observation interval; and

o 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
stream can occupy the physical medium.

The talker promises that these limits will not be exceeded. If the
talker 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 streams would serve no purpose. Those other
streams have their own dedicated resources, on the assumption that
all DetNet streams can use all of their resources over a long period
of time.

Note that there is no provision in DetNet for throttling streams; the
assumption is that a DetNet stream, 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 that 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) streams means that bridges and routers have to dedicate
buffer resources to specific streams or to classes of streams. 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.

4.5.2. Incomplete Networks

The presence in the network of relay systems that are not fully
capable of offering DetNet services complicates the ability of the
relay systems and/or controller to allocate resources, as extra
buffering, and thus extra latency, must be allocated at each point
that is downstream from the non-DetNet relay system for some DetNet
stream.

4.6. Data Flow Model through Systems

4.7. Queuing, Shaping, Scheduling, and Preemption

   For this reason, the IEEE 802.1 Time-Sensitive Networking Task Group
   has defined a 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 stream or class of streams.

4.8. Coexistence with normal traffic

   A DetNet network supports the dedication of at least 75% of the
   network bandwidth to DetNet streams. But, no matter how much is
   dedicated for DetNet streams, It is z goal of DetNet to not interfere
   excessively with existing QoS schemes. It is also important that
   non-DetNet traffic not disrupt the DetNet stream, of course (see
   Section 4.9 and Section 6). For these reasons:

   o Bandwidth (transmission opportunities) not utilized by a DetNet
      stream are available to non-DetNet packets (though not to other
      DetNet streams).

   o DetNet streams can be shaped, in order to ensure that the highest-
      priority non-DetNet packet also is ensured a maximum latency.

   o When transmission opportunities for DetNet streams are scheduled
      in detail, then the algorithm constructing the schedule should
      leave sufficient opportunities for non-DetNet packets to satisfy
      the needs of the uses of the network.

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

4.9. 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 disabling the offending packet,
   shutting down the offending DetNet stream, or shutting down the
   offending interface.

   It is also essential that filters and service remarking be employed
   to prevent non-DetNet packets from 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 systemd will have zero
impact on well-behaved DetNet streams, except of course, for the
receiving interface(s) immediately downstream of the misbehaving
device.

4.10. Protocol Stack Model

   This section will be further developed. See [IEEE802.1CB], Annex C,
   for a description of the protocol stack. This is very much a work in
   progress, not a standard. See also [IEEE802.1Qcc].

4.11. Advertising resources, capabilities and adjacencies

4.12. Provisioning model

4.12.1. Centralized Path Computation and Installation

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

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

   o 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.

4.12.2. Distributed Path Setup

   Whether a distributed alternative without a PCE can be valuable
   should be studied as well. Such an alternative could for instance
   inherit from the Resource ReSerVation Protocol [RFC5127] (RSVP)
   flows.

   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-2011] Clause 35 describes SRP, a peer-to-peer
   protocol for Layer-2 roughly analogous to RSVP. Almost complete is
   [IEEE802.1Qca], which defines how ISIS can provide multiple disjoint
   paths or distribution trees. Also in progress is [IEEE802.1Qcc],
   which expands the capabilities of SRP.

5. Related IETF work

5.1. Deterministic PHB

   [I-D.svshah-tsvwg-deterministic-forwarding] defines a Differentiated
   Services Per-Hop-Behavior (PHB) Group called Deterministic Forwarding
   (DF). The document describes the purpose and semantics of this PHB.
   It also describes creation and forwarding treatment of the service
   class. The document also describes how the code-point can be mapped
   into one of the aggregated Diffserv service classes [RFC5127].

5.2. 6TiSCH

   Industrial process control already leverages deterministic wireless
   Low power and Lossy Networks (LLNs) to interconnect critical
   resource-constrained devices and form wireless mesh networks, with
   standards such as [ISA100.11a] and [WirelessHART].

   These standards rely on variations of the [IEEE802154e] timeSlotted
   Channel Hopping (TSCH) [I-D.ietf-6tisch-tsch] Medium Access Control
   (MAC), and a form of centralized Path Computation Element (PCE), to
   deliver deterministic capabilities.

   The TSCH MAC benefits include high reliability against interference,
   low power consumption on characterized streams, and Traffic
   Engineering capabilities. Typical applications are open and closed
   control loops, as well as supervisory control streams and management.

   The 6TiSCH Working Group focuses only on the TSCH mode of the IEEE
   802.15.4e standard. The WG currently defines a framework for
   managing the TSCH schedule. Future work will standardize
   deterministic operations over so-called tracks as described in
   [I-D.ietf-6tisch-architecture]. Tracks are an instance of a
   deterministic path, and the DetNet work is a prerequisite to specify
   track operations and serve process control applications.

   [RFC5673] and [I-D.ietf-roll-rpl-industrial-applicability] section
   2.1.3. and next discusses application-layer paradigms, such as
   Source-sink (SS) that is a Multipeer to Multipeer (MP2MP) model that
   is primarily used for alarms and alerts, Publish-subscribe (PS, or
   pub/sub) that is typically used for sensor data, as well as Peer-to-
   peer (P2P) and Peer-to-multipeer (P2MP) communications. Additional
   considerations on Duocast and its N-cast generalization are also
   provided for improved reliability.

6. 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 coused by software failures. Because there is
   essentiall 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 4.9.

   Security must cover:

   o the protection of the signaling protocol

   o the authentication and authorization of the controlling systems

   o the identification and shaping of the streams

7. IANA Considerations

   This document does not require an action from IANA.

8. 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, Wilfried Steiner,
   Marcel Kiessling, Karl Weber, Ethan Grossman and Pat Thaler, for
   their various contribution with this work.

9. 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.

10. 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.finn-detnet-problem-statement]
              Finn, N. and P. Thubert, "Deterministic Networking Problem
              Statement", draft-finn-detnet-problem-statement-01 (work
              in progress), October 2014.

   [I-D.ietf-6tisch-architecture]
              Thubert, P., Watteyne, T., Struik, R., and M. Richardson,
              "An Architecture for IPv6 over the TSCH mode of IEEE
              802.15.4e", draft-ietf-6tisch-architecture-06 (work in
              progress), March 2015.

   [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-05 (work in
              progress), January 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.svshah-tsvwg-deterministic-forwarding]
             Shah, S. and P. Thubert, "Deterministic Forwarding PHB",
             draft-svshah-tsvwg-deterministic-forwarding-03 (work in
             progress), March 2015.

   [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, "Seamless Redundancy (IEEE Draft P802.1CB)", 2015,
             <http://www.ieee802.org/1/files/private/cb-drafts/>.

   [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>.

   [IEEE802.1Qca]
             IEEE, "Path Control and Reservation", 2015,
             <http://www.ieee802.org/1/files/private/ca-drafts/>.

   [IEEE802.1Qcc]
             IEEE, "Stream Reservation Protocol (SRP) Enhancements and
             Performance Improvements", 2015,
             <http://www.ieee802.org/1/files/private/cc-drafts/>.

   [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", 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>.

[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, B., Zhang, L., Berson, S., Herzog, S., and S.
          Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1
          Functional Specification", RFC 2205, September 1997.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V.,
          and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP
          Tunnels", RFC 3209, December 2001.

[RFC4655] Farrel, A., Vasseur, J., and J. Ash, "A Path Computation
          Element (PCE)-Based Architecture", RFC 4655, August 2006.

[RFC5127] Chan, K., Babiarz, J., and F. Baker, "Aggregation of
          Diffserv Service Classes", RFC 5127, February 2008.

[RFC5673] Pister, K., Thubert, P., Dwars, S., and T. Phinney,
          "Industrial Routing Requirements in Low-Power and Lossy
          Networks", RFC 5673, October 2009.

   [RFC7384] Mizrahi, T., "Security Requirements of Time Protocols in
              Packet Switched Networks", RFC 7384, October 2014.

   [RFC7426] Haleplidis, E., Pentikousis, K., Denazis, S., Hadi Salim,
              J., Meyer, D., and O. Koufopavlou, "Software-Defined
              Networking (SDN): Layers and Architecture Terminology",
              RFC 7426, January 2015.

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

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

Authors' Addresses

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

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

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

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

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

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

Deterministic Networking Problem Statement
draft-finn-detnet-problem-statement-03

Abstract

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

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

   Operational Technology (OT) refers to industrial networks that are
   specifically deployed in order to monitor production systems and
   support control loops and movement detection operations for process
   control (i.e., continuous manufacturing) and factory automation
   (i.e., discrete manufacturing), as well as protection systems used in
   power distribution automation (the SmartGrid). Due to its different
   goals, OT has evolved in parallel but in a manner that is radically
   different from Information Technology/Information and Communications
   Technology (IT/ICT), focusing on highly secure, reliable and
   deterministic networks, with limited scalability over a bounded and
   closed area.

   In OT environments, deterministic networks are characterized as
   providing a guaranteed bandwidth with extremely low packet loss
   rates, bounded latency, and low jitter.

   The convergence of IT and OT technologies, also called the Industrial
   Internet, represents a major evolution for both sides. For IT, it
   means a new level of Quality of Service whereby the transfer of
   packets is completely controlled and repeatable, different flows are
   perfectly isolated from one another, and packet loss and system

downtimes are reduced drastically; for OT, it means sharing IT
resources between deterministic and stochastic flows in order to
retrieve vasts amounts of so-far unmeasured data and enable
additional optimizations.

The work has already started; in particular, the industrial
automation space has been developing a number of Ethernet-based
replacements for existing digital control systems (DCS), often not
packet-based (fieldbus technologies). These replacements are meant
to provide similar behavior as the incumbent protocols, and their
common focus is to transport a fully characterized flow over a well-
controlled environment (i.e., a factory floor), with a bounded
latency, extraordinarily low frame loss, and a very narrow jitter.
Examples of such protocols include PROFINET, ODVA Ethernet/IP, and
EtherCAT.

As an example, Industrial Automation segregates the network along the
broad lines of the Purdue Enterprise Reference Architecture (PERA),
using different technologies at each level, and public
infrastructures such as the power distribution grid require
deterministic properties over the Wide Area. To fully serve an
industrial application between a wireless sensor and a virtualized
control system operating from the carpeted floor, a deterministic
path may span, for instance, across a (limited) number of 802.1
bridges and then a (limited) number of IP routers. In that example,
the IEEE802.1 bridges may be operating at Layer-2 over Ethernet
whereas the IP routers may be 6TiSCH [TiSCH] nodes operating at
Layer-2 and/or Layer-3 over the IEEE802.15.4 MAC.

In parallel, the need for determinism in professional and home audio/
video markets drove the formation of the Audio/Video Bridging (AVB)
standards efforts in IEEE 802.1. With the demand for connectivity
and multimedia in transportation, AVB is being evaluated for
application in vehicle head units, rear seat entertainment modules,
amplifiers, camera modules, and engine control systems. Automotive
AVB networks share the OT requirements for deterministic networks
characteristics.

Other instances of in-vehicle deterministic networks have arisen as
well for control networks in cars, trains and buses, as well as
avionics, with, for instance, the mission-critical "Avionics Full-
Duplex Switched Ethernet" (AFDX) that was designed as part of the
ARINC 664 standards. Existing automotive control networks such as
the LIN, CAN and FlexRay standards were not designed to cover the
increasing demands in terms of bandwidth and scalability that we see
with various kinds of Driver Assistance Systems (DAS); it results
that new multiplexing technologies based on Ethernet are now getting
traction.

Other industries where strong needs for deterministic networks are now emerging include: radio access networks [I-D.korhonen-detnet-telreq], the SmartGrid [I-D.wetterwald-detnet-utilities-reqs], and ProAudio networks [I-D.gunther-detnet-proaudio-req].

This wider application scope for deterministic networks has led to the IEEE802.1 AVB Task Group becoming the Time-Sensitive Networking (TSN) Task Group (TG) [IEEE802.1TSNTG], additionally covering industrial and vehicular applications.

The networks in consideration are now extending beyond the LAN boundaries and require secure deterministic forwarding and connectivity over a mixed Layer-2/Layer-3 network. 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 such as PCE [PCE], TEAS [TEAS], CCAMP [CCAMP] and MPLS [MPLS], 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 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],[HSR-PRP], 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 (see Section 5.2) that enable each bridge or router to compute the exact number of buffers to be allocated for each flow or class of flows. The present authors

assume that these techniques will be used by the DetNet Working
Group.

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. Related IETF work

4.1. Deterministic PHB

[I-D.svshah-tsvwg-deterministic-forwarding] defines a Differentiated
Services Per-Hop-Behavior (PHB) Group called Deterministic Forwarding
(DF). The document describes the purpose and semantics of this PHB.
It also describes creation and forwarding treatment of the service
class, and how the code-point can be mapped into one of the
aggregated Diffserv service classes [RFC5127].

4.2. 6TiSCH

Industrial process control already leverages deterministic wireless
Low power and Lossy Networks (LLNs) to interconnect critical
resource-constrained devices and form wireless mesh networks, with
standards such as [ISA100.11a] and [WirelessHART].

These standards rely on variations of the [IEEE802154] timeSlotted
Channel Hopping (TSCH) [RFC7554] Medium Access Control (MAC), and a
form of centralized Path Computation Element (PCE), to deliver
deterministic capabilities.

The TSCH MAC benefits include high reliability against interference,
low power consumption on characterized flows, and Traffic Engineering
capabilities. Typical applications are open and closed control
loops, as well as supervisory control flows and management.

The 6TiSCH Working Group focuses only on the TSCH mode of the
IEEE802.15.4e standard. The WG currently defines a framework for
managing the TSCH schedule. Future work will standardize
deterministic operations over so-called tracks as described in
[I-D.ietf-6tisch-architecture]. Tracks are an instance of a
deterministic path, and the DetNet work is a prerequisite to specify
track operations and serve process control applications. The
dependencies that 6TiSCH has on PCE and DetNet work are further
discussed in [I-D.thubert-6tisch-4detnet].

[RFC5673] and [I-D.ietf-roll-rpl-industrial-applicability] section
2.1.3 and next discuss application-layer paradigms, such as Source-
sink (SS) that is a Multipeer to Multipeer (MP2MP) model that is
primarily used for alarms and alerts, Publish-subscribe (PS, or pub/
sub) that is typically used for sensor data, as well as Peer-to-peer
(P2P) and Peer-to-multipeer (P2MP) communications. Additional
considerations on Duocast and its N-cast generalization are also
provided for improved reliability.

5. Related work in other standards organizations

5.1. Bridged solutions

Completed and ongoing work in other standards bodies have, to date,
produced viable solutions, suitable for carrying IP traffic for a
subset of the applications of interest to DetNet, but only over
bridged networks, not through routers. Among these are:

o IEEE 802 Audio-Video Bridging [IEEE802.1BA-2011].

o IEEE 802 Time-Sensitive Networking (TSN) Task Group (TG)
   [IEEE802.1TSNTG]

o ISO/IEC HSR and PRP [HSR-PRP].

5.2. Queuing and shaping

A number of standards are completed or in progress in the IEEE 802.1
(bridging) and IEEE 802.3 (Ethernet) Working Groups related to the
queuing and transmission of Ethernet frames. Most of these standards
could be applied to non-Ethernet or non-802 media with equal
facility, and so will likely be of use to DetNet. See the DetNet
architecture draft [I-D.finn-detnet-architecture] for a detailed
list.

6. Problem Statement

6.1. DetNet architecture

   An architecture that defines the space in which the various parts of
   the DetNet solution operate is required. A start has been made with
   [I-D.finn-detnet-architecture]. The main consideration is to build
   on art that is deployed in existing OT networks.

   These networks are systematically designed around a central
   controller that has a God's view on the devices, their capabilities,
   and their links to neighbors. The controller gets requests to
   establish flows with certain Traffic Specifications, and programs the
   necessary resources in the network to support those flows.

   This design, referred to as Software Defined Networking (SDN),
   simplifies the computation and the setup of paths, and ensures a
   better view and an easier control of the network by an operator. To
   inherit from this art, it has been determined early in DetNet
   discussions that the work would initially focus on an SDN model as
   well.

   DetNet should thus produce the complete SDN architecture with
   describes at a high level the interaction and data models to:

   o report the topology and device capabilities to the central
      controller;

   o request a path setup for a new flow with particular
      characteristics over the service interface and control it through
      its life cycle;

   o signal the new path to the devices, modify it to cope with various
      events such as loss of a link, update it and tear it down;

   o expose the status of the path to the end devices (UNI interface)

   o provide additional reliability through redundancy, in particular
      with packet replication and elimination;

   o indicate the flows and packet sequences in-band with the flows;

   The related concepts are already laid out at the IETF with [RFC7426],
   which introduces the following elements:

SDN Layers and Architecture Terminology per RFC 7426

```
          o------------------------------o
          |                              |
          | +-------------+  +----------+ |
          | | Application |  | Service  | |
          | +-------------+  +----------+ |
          |         Application Plane     |
          o---------------Y--------------o
                          |
*-------------------------------Y---------------------------------*
|           Network Services Abstraction Layer (NSAL)             |
*------Y----------------------------------------------Y-------*
       |                                              |
       |              Service Interface               |
       |                                              |
 o------Y----------------o       o--------------------Y------o
 |      |   Control Plane |       | Management Plane   |     |
 | +----Y----+  +-----+   |       | +-----+    +----Y----+   |
 | | Service |  | App |   |       | | App |    | Service |   |
 | +----Y----+  +--Y--+   |       | +--Y--+    +----Y----+   |
 |      |          |      |       |    |           |         |
 | *----Y----------Y----* |       | *---Y--------------Y----* |
 | | Control Abstraction | |       | | Management Abstraction | |
 | |    Layer (CAL)      | |       | |     Layer (MAL)      | |
 | *----------Y----------* |       | *----------Y-------------* |
 |            |            |       |            |              |
 o------------|-----------o       o------------|--------------o
              |                                 |
              | CP                              | MP
              | Southbound                      | Southbound
              | Interface                       | Interface
              |                                 |
*------------Y----------------------------------Y--------------*
|           Device and resource Abstraction Layer (DAL)         |
*------------Y----------------------------------Y--------------*
 |           |                      |           |              |
 |    o-------Y----------o  +-----+  o--------Y----------o      |
 |    | Forwarding Plane |  | App |  | Operational Plane |      |
 |    o-----------------o  +-----+  o------------------o       |
 |                   Network Device                            |
 +-----------------------------------------------------------+
```
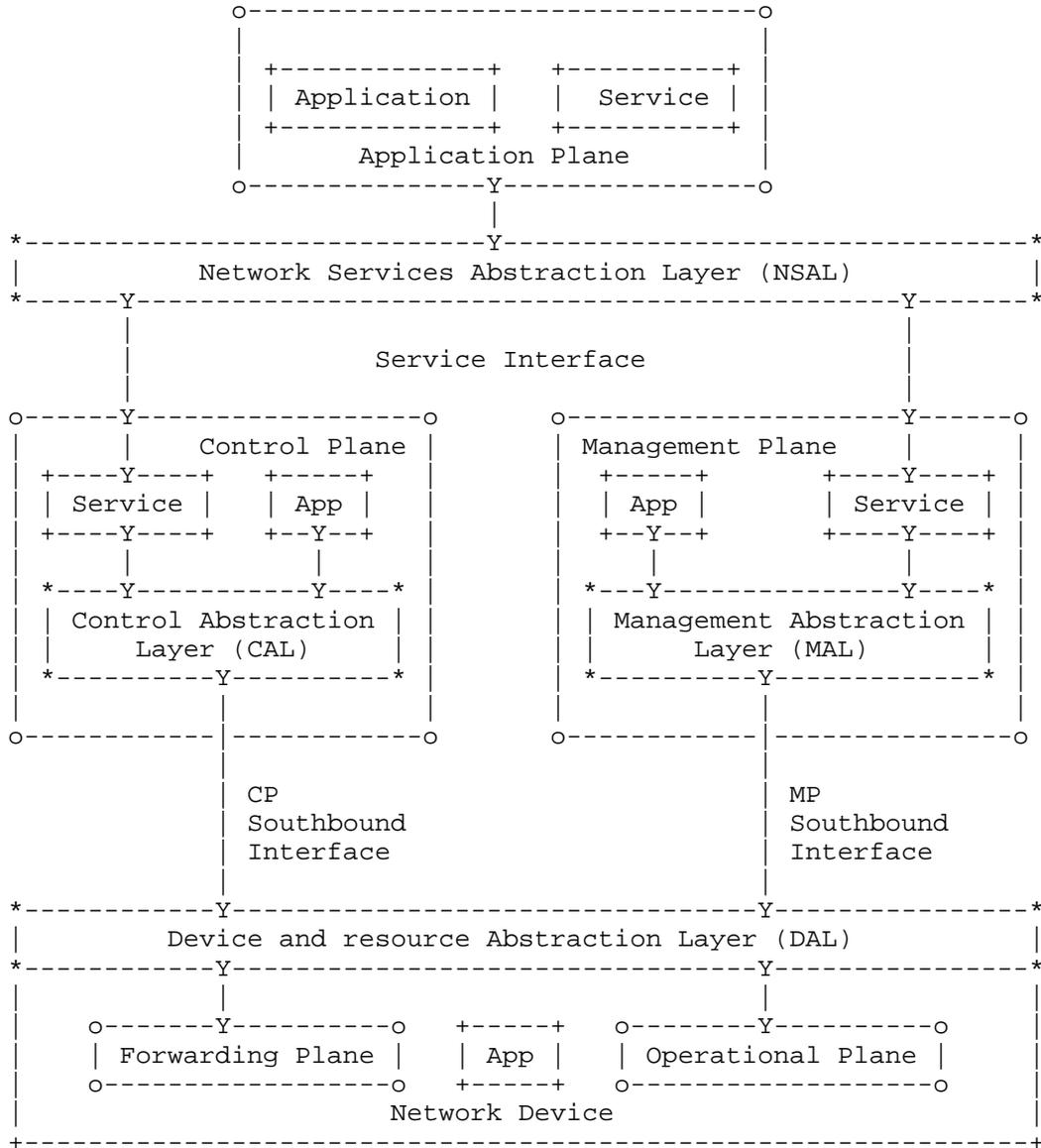
Figure 1

6.2. 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.

6.3. 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:

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

   o 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.

6.4. 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 [RFC5127] (RSVP) flows. But
   the focus of the work should be to deliver the centralized approach
   first.

6.5. 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.

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.

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 if 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:

o the protection of the signaling protocol

o the authentication and authorization of the controlling nodes

o the identification and shaping of the flows

o the isolation of flows from leakage and other influences from any activity sharing physical resources.

8. IANA Considerations

   This document does not require an action from IANA.

9. 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, Wilfried Steiner, Marcel Kiessling, Karl Weber, Ethan Grossman and Pat Thaler, for their various contribution with this work.

10. References

10.1. Normative References

   [I-D.gunther-detnet-proaudio-req]
             Gunther, C. and E. Grossman, "Deterministic Networking
             Professional Audio Requirements", draft-gunther-detnet-
             proaudio-req-01 (work in progress), March 2015.

   [I-D.korhonen-detnet-telreq]
             Korhonen, J., "Deterministic networking for radio access
             networks", draft-korhonen-detnet-telreq-00 (work in
             progress), May 2015.

   [I-D.thubert-6tisch-4detnet]
             Thubert, P., "6TiSCH requirements for DetNet", draft-
             thubert-6tisch-4detnet-00 (work in progress), June 2015.

   [I-D.wetterwald-detnet-utilities-reqs]
             Wetterwald, P. and J. Raymond, "Deterministic Networking
             Uitilities requirements", draft-wetterwald-detnet-
             utilities-reqs-01 (work in progress), October 2014.

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
             Requirement Levels", BCP 14, RFC 2119, March 1997.

10.2. 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/>.

   [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.".

   [I-D.finn-detnet-architecture]
             Finn, N., Thubert, P., and M. Teener, "Deterministic
             Networking Architecture", draft-finn-detnet-
             architecture-01 (work in progress), March 2015.

   [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-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-03 (work in
             progress), March 2015.

   [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>.

   [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/>.

   [RFC2205] Braden, B., Zhang, L., Berson, S., Herzog, S., and S.
              Jamin, "Resource ReSerVation Protocol (RSVP) -- Version 1
              Functional Specification", RFC 2205, September 1997.

   [RFC5127] Chan, K., Babiarz, J., and F. Baker, "Aggregation of
              Diffserv Service Classes", RFC 5127, February 2008.

   [RFC5673] Pister, K., Thubert, P., Dwars, S., and T. Phinney,
              "Industrial Routing Requirements in Low-Power and Lossy
              Networks", RFC 5673, October 2009.

      [RFC7384] Mizrahi, T., "Security Requirements of Time Protocols in
                 Packet Switched Networks", RFC 7384, October 2014.

      [RFC7426] Haleplidis, E., Pentikousis, K., Denazis, S., Hadi Salim,
                 J., Meyer, D., and O. Koufopavlou, "Software-Defined
                 Networking (SDN): Layers and Architecture Terminology",
                 RFC 7426, January 2015.

      [RFC7554] Watteyne, T., 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,
                 May 2015.

      [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

Deterministic Networking Professional Audio Requirements
draft-gunther-detnet-proaudio-req-01

Abstract

   This draft documents the needs in the professional audio and video
   industry to establish multi-hop paths and optional redundant 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 a Layer 3 (IP) interface.

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 October 2, 2015.

Copyright Notice

the Trust Legal Provisions and are provided without warranty as
described in the Simplified BSD License.

Table of Contents

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. Requirements Language

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

3. 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.

3.1. 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.

3.2. 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.

3.2.1. 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.

4. Additional Stream Requirements

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

4.1. 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).

4.2. 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.

4.3. 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.

4.4. 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.

4.5. 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.

4.6. 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).

4.7. 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.

4.7.1. 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.

4.7.2. 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.

5. 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.

6. 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.

6.1. 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.

6.2. 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.

7. 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.

8. 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

9. IANA Considerations

   This memo includes no request to IANA.

10. References

10.1. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

10.2. Informative References

   [CONTENT_PROTECTION]
              Olsen, D., "1722a Content Protection", 2012,
              <http://grouper.ieee.org/groups/1722/contributions/2012/
              avtp_dolsen_1722a_content_protection.pdf>.

   [DCI]   Digital Cinema Initiatives, LLC, "DCI Specification,
              Version 1.2", 2012, <http://www.dcimovies.com/>.

   [ESPN_DC2]
              Daley, D., "ESPN's DC2 Scales AVB Large", 2014,
              <http://sportsvideo.org/main/blog/2014/06/
              espns-dc2-scales-avb-large>.

   [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>.

   [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>.

Authors' Addresses

   Craig Gunther (editor)
   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

Ethan Grossman (editor)
Dolby Laboratories, Inc.
100 Potrero Ave
San Francisco, CA 94103
USA

Phone: +1 415 645 4726
Email: ethan.grossman@dolby.com
URI: http://www.dolby.com

                 Deterministic networking for radio access networks
                      draft-korhonen-detnet-telreq-00

Abstract

   This document describes requirements for deterministic networking in
   cellular radio access and 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.

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 November 28, 2015.

Table of Contents

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.

2. Network architecture

Figure Figure 1 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 3 will discuss the PDV requirements
in more detail.

```
         Y (remote radios)
          \
     Y__   \.--.                      .--.          +------+
        \_(      `.       +---+      _(Back`.       | 3GPP |
    Y------( Front  )----|eNB|----(  Haul  )----| core |
       (  ` .Haul )      +---+    (  `  .  )  )     | netw |
      /'--(___.-'         \         `--(___.-'      +------+
    Y_/      /             \.--.                \
       Y_/              _( Mid`.          \
                       (   Haul )          \
                       ( `  .  )  )          \
                        `--(___.-'_____+---+     (small cells)
                            \           |SCe|__Y
                          +---+         +---+
                      Y__|eNB|__Y
                          +---+
                      Y_/   \_Y ("local" radios)
```
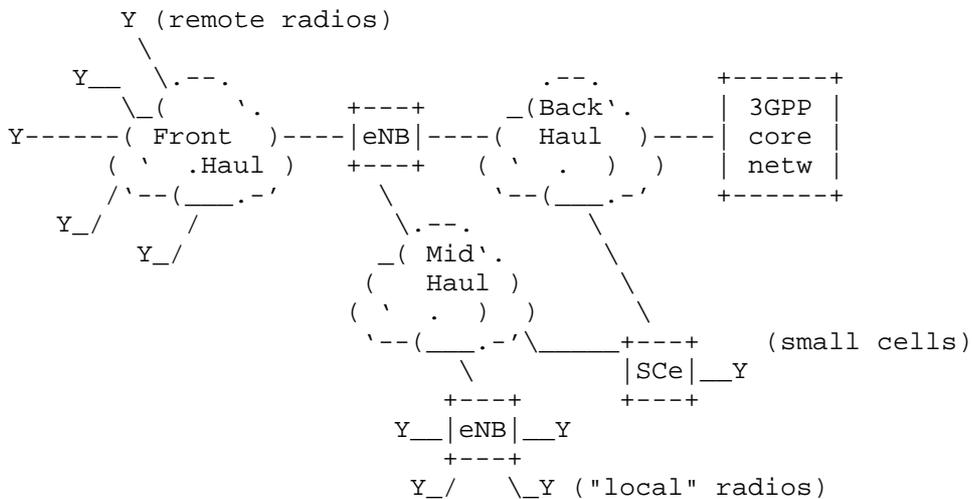
       Figure 1: Generic 3GPP-based cellular network architecture with
                       Front/Mid/Backhaul networks

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.

4. Time-sensitive stream requirements

In addition to the time synchronization requirements listed in
Section Section 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.

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.

6. IANA Considerations

   This document has no IANA considerations.

7. Acknowledgements

    The author(s) ACK and NACK.

8. Informative References

    [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>.

    [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>.

    [I-D.finn-detnet-architecture]
                Finn, N., Thubert, P., and M. Teener, "Deterministic
                Networking Architecture", draft-finn-detnet-
                architecture-01 (work in progress), March 2015.

    [I-D.finn-detnet-problem-statement]
                Finn, N. and P. Thubert, "Deterministic Networking Problem
                Statement", draft-finn-detnet-problem-statement-01 (work
                in progress), October 2014.

    [I-D.ietf-tictoc-1588overmpls]
                Davari, S., Oren, A., Bhatia, M., Roberts, P., and L.
                Montini, "Transporting Timing messages over MPLS
                Networks", draft-ietf-tictoc-1588overmpls-06 (work in
                progress), April 2014.

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

   [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>.

   [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>.

   [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>.

   [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>.

   [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
             Label Switching Architecture", RFC 3031, January 2001.

   [RFC3985] Bryant, S. and P. Pate, "Pseudo Wire Emulation Edge-to-
             Edge (PWE3) Architecture", RFC 3985, March 2005.

[RFC4553]   Vainshtein, A. and YJ. Stein, "Structure-Agnostic Time
            Division Multiplexing (TDM) over Packet (SAToP)", RFC
            4553, June 2006.

[RFC5086]   Vainshtein, A., 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, December 2007.

[RFC5087]   Stein, Y(J)., Shashoua, R., Insler, R., and M. Anavi,
            "Time Division Multiplexing over IP (TDMoIP)", RFC 5087,
            December 2007.

[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>.

[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>.

Author's Address

   Jouni Korhonen
   Broadcom Corporation
   3151 Zanker Road
   San Jose, CA 95134
   USA

   Email: jouni.nospam@gmail.com

                    6TiSCH requirements for DetNet
                    draft-thubert-6tisch-4detnet-01

Abstract

   This document builds on the 6TiSCH architecture that defines, among
   others, mechanisms to establish and maintain deterministic routing
   and scheduling in a centralized fashion. The document details
   dependencies on DetNet and PCE controller to express topologies and
   capabilities, as well as abstract state that the controller must be
   able to program into the network devices to enable deterministic
   forwarding operations.

Requirements Language

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

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 December 13, 2015.

Copyright Notice

Table of Contents

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)
[I-D.ietf-6tisch-tsch] 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.

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.

3. 6TiSCH Overview

    The scope of the present work is a subnet that, in its basic
    configuration, is made of a TSCH [I-D.ietf-6tisch-tsch] MAC Low Power
    Lossy Network (LLN).

```
              ---+-------- ............ ------------
                 |       External Network     |
                 |                        +-----+
              +-----+                     | NME |
              |     | LLN Border          |     |
              |     | router              +-----+
              +-----+
                 o     o    o
              o     o   o        o
                o   o LLN   o    o       o
                  o   o   o        o
                        o
```
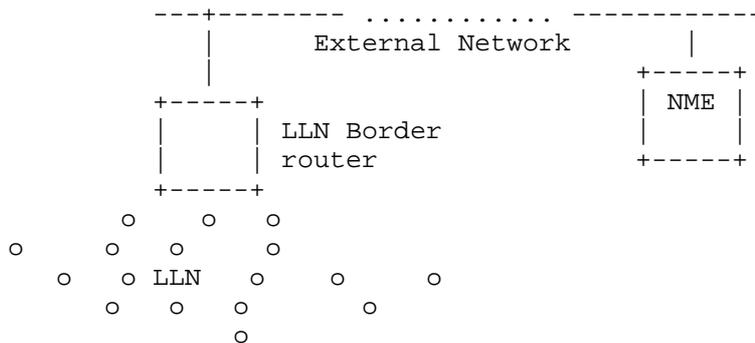
              Figure 1: Basic Configuration of a 6TiSCH Network

    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.

```
              ---+-------- ........... ------------
                 |    External Network      |
                 |                       +-----+
                 |            +-----+     | NME |
              +-----+         | +-----+   |     |
              |     | Router  | | PCE |   +-----+
              |     |         +--|     |
              +-----+         +-----+
                 |               |
                 | Subnet Backbone |
              +------------------+----------------+
                 |               |              |
              +-----+         +-----+        +-----+
              |     | Backbone  |     | Backbone  |     | Backbone
          o   |     | router    |     | router    |     | router
              +-----+         +-----+        +-----+
          o                o               o          o   o
              o   o   o       o   o   o   o    o   o   o   o
          o            o        o LLN    o     o      o     o
              o   o   o     o      o o    o   o   o   o   o   o
```
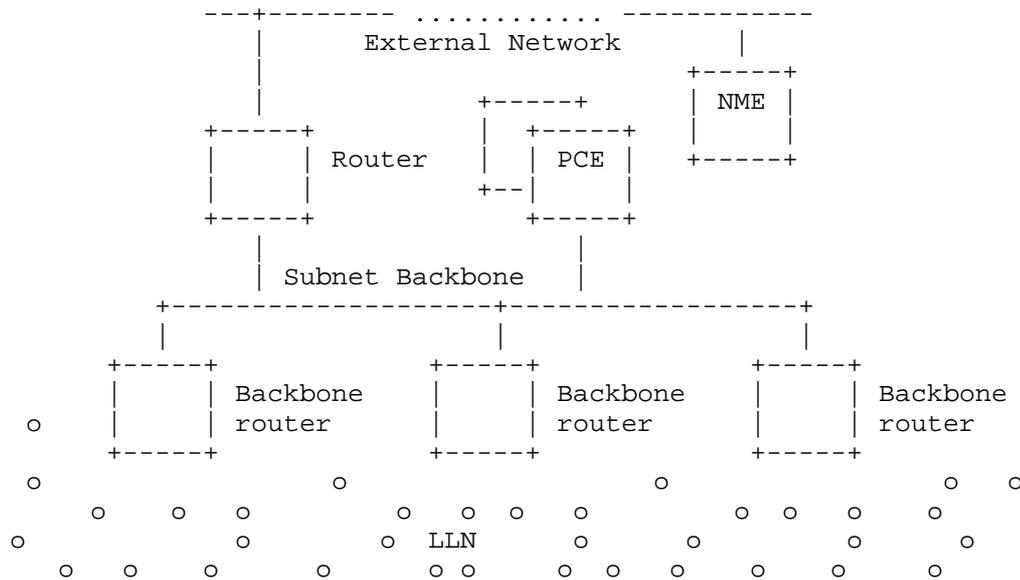
                Figure 2: Extended Configuration of a 6TiSCH Network

   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.

```
                        +-----+
                        | IoT |
                        | G/W |
                        +-----+
                          ^  <---- Elimination
                          | |
        Track branch      | |
                 +-------+ +--------+ Subnet Backbone
                     |             |
           +--|--+              +--|--+
           |  |  | Backbone     |  |  | Backbone
      o    |  |  | router       |  |  | router
           +--/--+              +--|--+
       o    /    o      o---o----/       o
         o    o---o--/    o       o   o   o    o
      o     \  /     o                o   LLN      o
         o   v  <---- Replication
              o
```
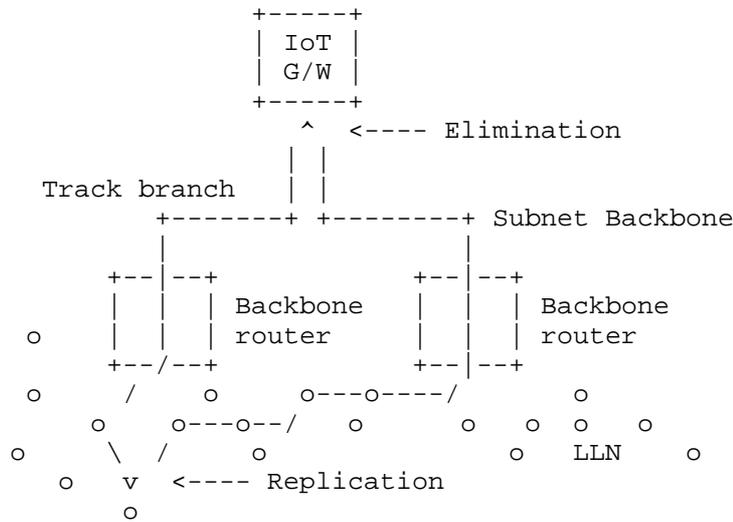
Figure 3: End-to-End deterministic Track

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.

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.

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 fond in section "SlotFrames and Priorities" of
   [I-D.ietf-6tisch-architecture]

   IEEE802.15.4 TSCH avoids contention on the medium by formatting time
   and frequencies in cells of transmission of equal duration. In order
   to describe that formatting of time and frequencies, the 6TiSCH
   architecture defines a global concept that is called a Channel
   Distribution and Usage (CDU) matrix; a CDU matrix is a matrix of
   cells with an height equal to the number of available channels
   (indexed by ChannelOffsets) and a width (in timeSlots) that is the
   period of the network scheduling operation (indexed by slotOffsets)
   for that CDU matrix. The size of a cell is a timeSlot duration, and
   values of 10 to 15 milliseconds are typical in 802.15.4 TSCH to
   accommodate for the transmission of a frame and an ack, including the
   security validation on the receive side which may take up to a few
   milliseconds on some device architecture.

   The frequency used by a cell in the matrix rotates in a pseudo-random
   fashion, from an initial position at an epoch time, as the matrix
   iterates over and over.

   A CDU matrix is computed by the PCE, but unallocated timeSlots may be
   used opportunistically by the nodes for classical best effort IP
   traffic. The PCE has precedence in the allocation in case of a
   conflict.

   In a given network, there might be multiple CDU matrices that operate
   with different width, so they have different durations and represent

different periodic operations. It is recommended that all CDU
matrices in a 6TiSCH domain operate with the same cell duration and
are aligned, so as to reduce the chances of interferences from
slotted-aloha operations. The PCE MUST compute the CDU matrices and
shared that knowledge with all the nodes. The matrices are used in
particular to define slotFrames.

A slotFrame is a MAC-level abstraction that is common to all nodes
and contains a series of timeSlots of equal length and precedence.
It is characterized by a slotFrame_ID, and a slotFrame_size. A
slotFrame aligns to a CDU matrix for its parameters, such as number
and duration of timeSlots.

Multiple slotFrames can coexist in a node schedule, i.e., a node can
have multiple activities scheduled in different slotFrames, based on
the precedence of the 6TiSCH topologies. The slotFrames may be
aligned to different CDU matrices and thus have different width.
There is typically one slotFrame for scheduled traffic that has the
highest precedence and one or more slotFrame(s) for RPL traffic. The
timeSlots in the slotFrame are indexed by the SlotOffset; the first
cell is at SlotOffset 0.

The 6TiSCH architecture introduces the concept of chunks
([I-D.ietf-6tisch-terminology]) to operate such spectrum distribution
for a whole group of cells at a time. The CDU matrix is formatted
into a set of chunks, each of them identified uniquely by a chunk-ID.
The PCE MUST compute the partitioning of CDU matrices into chunks and
shared that knowledge with all the nodes in a 6TiSCH network.

```
              +-----+-----+-----+-----+-----+-----+-----+     +-----+
chan.Off. 0   |chnkA|chnkP|chnk7|chnkO|chnk2|chnkK|chnk1| ... |chnkZ|
              +-----+-----+-----+-----+-----+-----+-----+     +-----+
chan.Off. 1   |chnkB|chnkQ|chnkA|chnkP|chnk3|chnkL|chnk2| ... |chnk1|
              +-----+-----+-----+-----+-----+-----+-----+     +-----+
                 ...
              +-----+-----+-----+-----+-----+-----+-----+     +-----+
chan.Off. 15  |chnkO|chnk6|chnkN|chnk1|chnkJ|chnkZ|chnkI| ... |chnkG|
              +-----+-----+-----+-----+-----+-----+-----+     +-----+
                 0     1     2     3     4     5     6          M
```
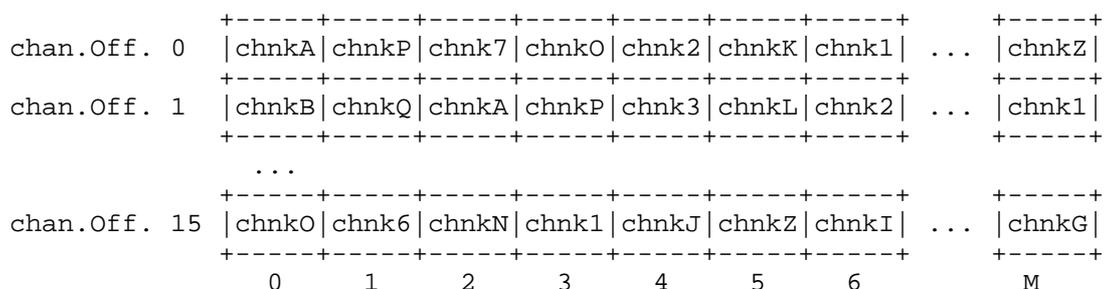
Figure 4: CDU matrix Partitioning in Chunks

The appropriation of a chunk can be requested explicitly by the PCE
to any node. After a successful appropriation, the PCE owns the
cells in that chunk, and may use them as hard cells to set up Tracks.

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.

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 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.

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.

```
                           |                              ^
 +--------------+          |                              |
 |     IPv6     |          |                              |
 +--------------+          |                              |
 |  6LoWPAN HC  |          |                              |
 +--------------+  ingress                          egress
 |     6top     |  sets      +----+          +----+  restores
 +--------------+  dmac to   |    |          |    |  dmac to
 |   TSCH MAC   |  brdcst    |    |          |    |   self
 +--------------+  |         |    |          |    |   |
 |   LLN PHY    |  +-------+ +--...-----+ +-------+
 +--------------+
```
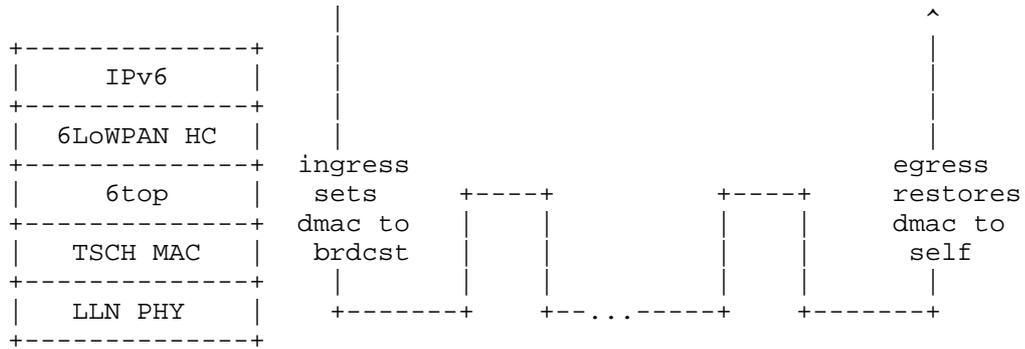
                Track Forwarding, Transport Mode

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.

```
+--------------+
|     IPv6     |
+--------------+
| 6LoWPAN HC   |                set            restore
+--------------+              +dmac+          +dmac+
|    6top      |            to|brdcst       to|nexthop
+--------------+            to|  |  |       to|  |  |
|   TSCH MAC   |              |  |  |          |  |  |
+--------------+              |  |  |          |  |  |
|   LLN PHY    |   +-------+  +--...-----+  +-------+
+--------------+   |  ingress               egress  |
                   |                                 |
+--------------+   |                                 |
|   LLN PHY    |   |                                 |
+--------------+   |                                 |
|   TSCH MAC   |   |                                 |
+--------------+   | dmac =                          | dmac =
|ISA100/WiHART |   | nexthop                         v nexthop
+--------------+
```
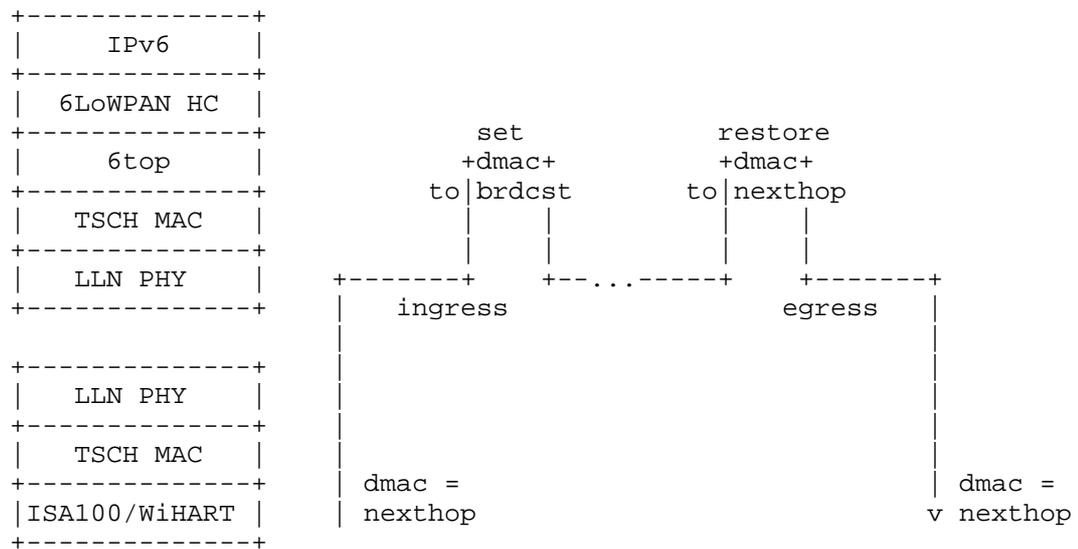
Figure 5: Track Forwarding, Tunnel Mode

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.

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.

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.

                        Stream Management Entity

                      Operational System and HMI

        -+-+-+-+-+-+-+ Northbound -+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

                PCE   PCE   PCE   PCE

        -+-+-+-+-+-+-+ Southbound -+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-

              --- 6TiSCH------6TiSCH------6TiSCH------6TiSCH--
        6TiSCH /   Device   Device   Device   Device   \
        Device-   - 6TiSCH
             \   6TiSCH   6TiSCH   6TiSCH   6TiSCH   / Device
              ----Device------Device------Device------Device--
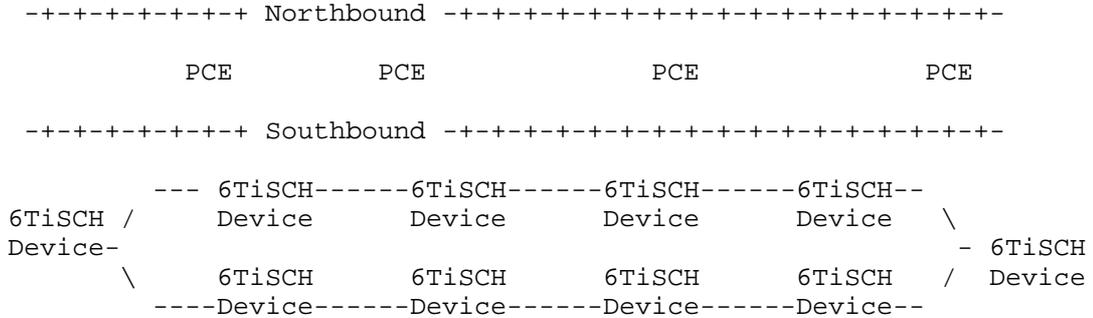
                                Figure 6

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.

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.

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.

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].

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. IANA Considerations

   This specification does not require IANA action.

6. 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.

7. 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.

8. References

8.1. Normative References

   [I-D.ietf-6tisch-6top-interface]
              Wang, Q., Vilajosana, X., and T. Watteyne, "6TiSCH
              Operation Sublayer (6top) Interface", draft-ietf-6tisch-
              6top-interface-03 (work in progress), March 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-04 (work in
              progress), March 2015.

   [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.

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

   [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6
              (IPv6) Specification", RFC 2460, December 1998.

8.2. Informative References

   [I-D.finn-detnet-architecture]
              Finn, N., Thubert, P., and M. Teener, "Deterministic
              Networking Architecture", draft-finn-detnet-
              architecture-01 (work in progress), March 2015.

   [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.svshah-tsvwg-deterministic-forwarding]
          Shah, S. and P. Thubert, "Deterministic Forwarding PHB",
          draft-svshah-tsvwg-deterministic-forwarding-03 (work in
          progress), March 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., Vilajosana, X., and T. Watteyne, "6TiSCH
          Operation Sublayer (6top)", draft-wang-6tisch-6top-
          sublayer-01 (work in progress), July 2014.

[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, December
          1998.

[RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V.,
          and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP
          Tunnels", RFC 3209, December 2001.

[RFC3444] Pras, A. and J. Schoenwaelder, "On the Difference between
          Information Models and Data Models", RFC 3444, January
          2003.

[RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)",
          RFC 3972, March 2005.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing
          Architecture", RFC 4291, February 2006.

[RFC4903] Thaler, D., "Multi-Link Subnet Issues", RFC 4903, June
          2007.

[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, August 2007.

   [RFC6282] Hui, J. and P. Thubert, "Compression Format for IPv6
              Datagrams over IEEE 802.15.4-Based Networks", RFC 6282,
              September 2011.

   [RFC6550] Winter, T., Thubert, P., 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, March 2012.

   [RFC6551] Vasseur, JP., Kim, M., Pister, K., Dejean, N., and D.
              Barthel, "Routing Metrics Used for Path Calculation in
              Low-Power and Lossy Networks", RFC 6551, March 2012.

   [RFC6775] Shelby, Z., Chakrabarti, S., Nordmark, E., and C. Bormann,
              "Neighbor Discovery Optimization for IPv6 over Low-Power
              Wireless Personal Area Networks (6LoWPANs)", RFC 6775,
              November 2012.

8.3. Other Informative References

   [ACE]   IETF, "Authentication and Authorization for Constrained
              Environments", <https://datatracker.ietf.org/doc/charter-
              ietf-ace/>.

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

   [DICE]   IETF, "DTLS In Constrained Environments",
              <https://datatracker.ietf.org/doc/charter-ietf-dice/>.

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

   [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.

    [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>.

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

    [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.

Author's Address

    Pascal Thubert (editor)
    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

detnet   P. Wetterwald
Internet-Draft   Cisco
Intended status: Informational   J. Raymond
Expires: January 1, 2016   Hydro-Quebec
                                                           June 30, 2015

               Deterministic Networking Uitilities requirements
                    draft-wetterwald-detnet-utilities-reqs-02

Abstract

   This paper documents the needs in Smart Grid industry 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 January 1, 2016.

1. Introduction

   [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.

2. Requirements Language

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

3. Overview

   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.

4. 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.

4.1. General Telecommunications Requirements

   o IP Connectivity everywhere

   o Monitoring services everywhere and from different remote centers

   o Move services to a virtual data center

   o Unify access to applications / information from the corporate
      network

   o Unify services

   o Unified Communications Solutions

   o Mix of fiber and microwave technologies - obsolescence of SONET/
      SDH or TDM

   o Standardize grid telecommunications protocol to opened standard to
      ensure interoperability

   o Reliable Telecommunications for Transmission and Distribution
      Substations

    o IEEE 1588 time synchronization Client / Server Capabilities

    o Integration of Multicast Design

    o QoS Requirements Mapping

    o Enable Future Network Expansion

    o Substation Network Resilience

    o Fast Convergence Design

    o Scalable Headend Design

    o Define Service Level Agreements (SLA) and Enable SLA Monitoring

    o Integration of 3G/4G Technologies and future technologies

    o Ethernet Connectivity for Station Bus Architecture

    o Ethernet Connectivity for Process Bus Architecture

    o Protection, teleprotection and PMU (Phaser Measurement Unit) on IP

4.1.1. 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:

    o Grid monitoring, control, and protection data

    o Non-control grid data (e.g. asset data for condition-based
       monitoring)

    o Physical safety and security data (e.g. voice and video)

    o Remote worker access to corporate applications (voice, maps,
       schematics, etc.)

    o Field area network backhaul for smart metering, and distribution
       grid management

o 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.

4.2. 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.

4.2.1. 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.

4.2.1.1. 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:

o 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.

4.2.1.1.1. 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.

4.2.1.1.2. 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:

o 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.

o 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.

o 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.

4.2.1.1.2.1. Other traffic characteristics

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

o 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.

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

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

4.2.1.1.2.2. 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  |     less than 50ms - hitless     |
|              failure        |                                  |
|                             |                                  |
|    performance management   |          Yes, Mandatory          |
|                             |                                  |
|          Redundancy         |               Yes                |
|                             |                                  |
|          Packet loss        |           0.1% to 1%             |
+---------------------------+----------------------------------+
```

Table 1: Teleprotection network requirements

4.2.1.2. 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     |          Attribute           |
|         Requirement          |                              |
+------------------------------+------------------------------+
|    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

4.2.1.3. 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

4.2.1.4. 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

4.2.1.5. 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

4.2.1.6. 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

4.2.1.7. 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

4.2.1.8. 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

4.2.2. Distribution use case

4.2.2.1. 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   |                 80 ms                   |
|       delay         |                                         |
|                     |                                         |
|   Asymetric delay   |                  No                     |
|      Required       |                                         |
|                     |                                         |
|   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    |                 Yes                     |
|      required       |                                         |
|                     |                                         |
|   Recovery time on  |       Depends on customer impact        |
|     Node failure    |                                         |
|                     |                                         |
|     performance     |             Yes, Mandatory              |
|      management     |                                         |
|                     |                                         |
|     Redundancy      |                 Yes                     |
|                     |                                         |
|     Packet loss     |                 0.1%                    |
+---------------------+-----------------------------------------+
```

                Table 9: FLISR Communication Requirements

4.2.3. Generation use case

4.2.3.1. 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) | Attribute |
|                  Requirement                 |              |
+----------------------------------------------+--------------+
|            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

4.3. 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.

4.4. 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.

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

6.1. 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:

o Network or transport errors (e.g. malformed packets or excessive
   latency) can cause protocol failure.

o 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.

   o Protocol commands may be available that are capable of restarting
      communications and otherwise interrupting processes.

   o Protocol commands may be available that are capable of clearing,
      erasing, or resetting diagnostic information such as counters and
      diagnostic registers.

   o Protocol commands may be available that are capable of requesting
      sensitive information about the controllers, their configurations,
      or other need-to-know information.

   o 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.

   o Protocol commands may be available that are capable of
      broadcasting messages to many devices at once (i.e. a potential
      DoS).

   o Protocol commands may be available to query the device network to
      obtain defined points and their values (i.e. a configuration
      scan).

   o Protocol commands may be available that will list all available
      function codes (i.e. a function scan).

   o 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.

6.2. 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:

   o Integration of distributed energy resources

   o Proliferation of digital devices to enable management, automation,
      protection, and control

   o Regulatory mandates to comply with standards for critical
      infrastructure protection

   o Migration to new systems for outage management, distribution
      automation, condition-based maintenance, load forecasting, and
      smart metering

   o 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:

   o Integrity : data cannot be altered undetectably

   o Authenticity : the telecommunications parties involved must be
      validated as genuine

   o Authorization : only requests and commands from the authorized
      users can be accepted by the system

   o 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.

7. Acknowledgements

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

   Pascal Thubert, CTAO Cisco

8. References

8.1. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119, March 1997.

8.2. Informative References

   [I-D.finn-detnet-problem-statement]
              Finn, N. and P. Thubert, "Deterministic Networking Problem
              Statement", draft-finn-detnet-problem-statement-03 (work
              in progress), June 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.

Authors' Addresses

   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

Network Working Group   Y. Zha
Internet Draft   Huawei Technologies
Intended status: Informational
Expires: January 2016

                                                        July 1, 2015

              Deterministic Networking Use Case in Mobile Network
                      draft-zha-detnet-use-case-00

Abstract

   This document describes some high level use cases and scenarios
   with requirements on delay sensitive and deterministic networking.
   Not only the telecom industry but also vertical industries have
   been investigated. In addition to the 5G networking, industrial
   automation, automotive industry, media and gaming industry are
   typical related industries believed to be representative for the
   technical requirements on ultra-fast and ultra-reliability
   communications.

Copyright Notice

Table of Contents

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 [5G]. 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].

2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT",
"SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in
this document are to be interpreted as described in [RFC2119]. In
this document, these words will appear with that interpretation
only when in ALL CAPS. Lower case uses of these words are not to
be interpreted as carrying [RFC2119] significance.

3. 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.

4. 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].

4.1. CoMP Architecture

```
                    +------------------------+
                    |          CoMP          |
                    +--+------------------+--+
                       |                  |
                 +----------+       +-----------+
                 |  Uplink  |       | Downlink  |
                 +-----+----+       +--------+--+
                       |                     |
               -----------------       --------------------
               |       |       |       |        |         |
         +---------+ +----+ +-----+ +-----------+ +-----+ +-----+
         |  Joint  | | CS | | DPS | |   Joint   | | CS/ | | DPS |
         |Reception| |    | |     | |Transmission| | CB  | |     |
         +---------+ +----+ +-----+ +-----------+ +-----+ +-----+
              |                           |
              |----------                 |------------
              |         |                 |           |
         +-----------+ +---------+   +----------+ +------------+
         |   Joint   | |  Soft   |   | Coherent | |    Non-    |
         |Equalization| |Combining|   |    JT    | | Coherent JT|
         +-----------+ +---------+   +----------+ +------------+
```
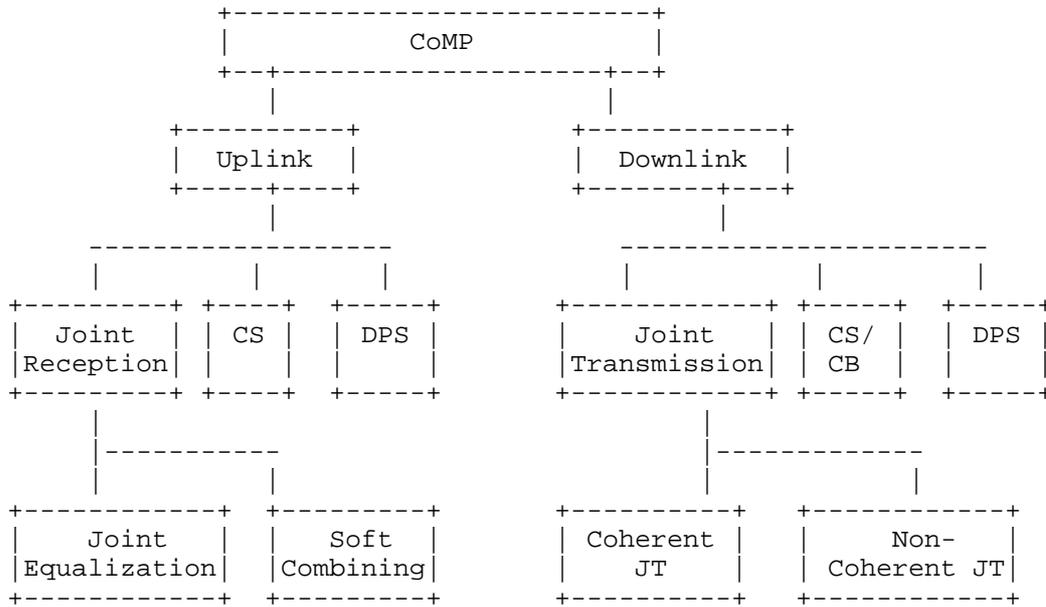
                Figure 1: Framework of CoMP Technology

   As shown in figure 1, 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.

4.2. 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.

5. 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.

6. 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. 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. Security Considerations

TBD

9. IANA Considerations

This document has no actions for IANA.

10. 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.

11. References

11.1. Normative References

[RFC2119] S. Bradner, "Key words for use in RFCs to Indicate
          Requirement Levels", BCP 14, RFC 2119, March 1997.

[RFC3393] C. Demichelis, "IP Packet Delay Variation Metric for IP
          Performance Metrics (IPPM) ", RFC 3393, Novermber 2002.

11.2. Informative References

[I-D.finn-detnet-problem-statement]

   Finn, N. and P. Thubert, "Deterministic Networking Problem
   Statement", draft-finn-detnet-problem-statement-01 (work in
   progress), October 2014.

   [I-D.finn-detnet-architecture]

   Finn, N., Thubert, P., and M. Teener, "Deterministic Networking
   Architecture", draft-finn-detnet-architecture-01 (work in
   progress), March 2015.

   [METIS] METIS Document Number: ICT-317669-METIS/D1.1, Scenarios,
   requirements and KPIs for 5G mobile and wireless system, April 29,
   2013. Available on line at: <https://www.metis2020.com/wp-
   content/uploads/deliverables/METIS_D1.1_v1.pdf>

   [5G] Ericsson white paper, "5G Radio Access, Challenges for 2020
   and Beyond." June 2013. Available at:
   <http://www.ericsson.com/res/docs/whitepapers/wp-5g.pdf>

   [CoMP] NGMN Alliance, "RAN EVOLUTION PROJECT COMP EVALUATION AND
   ENHANCEMENT ", MARCH 2015,
   <https://www.ngmn.org/uploads/media/NGMN_RANEV_D3_CoMP_Evaluation_
   and_Enhancement_v2.0.pdf>

   [LTE-Latency]Samuel Johnston, "LTE Latency: How does it compare to
   other technologies?" report of OpenSignal March 10, 2014.
   <http://opensignal.com/blog/2014/03/10/lte-latency-how-does-it-
   compare-to-other-technologies/>

   [EA12] P. C. Evans, M. Annunziata, "Industrial Internet: Pushing the
   Boundaries of Minds and Machines", General Electric White paper,
   November 2012.

   [UHD-video] Petr Holub, "Ultra-High Definition Videos and Their
   Applications over the Network", The 7th International Symposium on
   VICTORIES Project, OCTOBER 8, 2014. <http://www.aist-
   victories.org/jp/7th_sympo_ws/PetrHolub_presentation.pdf>

Authors' Addresses

    Yiyong Zha
    Huawei Technologies
    Email: zhayiyong@huawei.com