Deterministic Latency Network Use Cases
draft-liu-dln-use-cases-00

Abstract

   This draft documents low latency requirements in several diverse
   industries including virtual reality, electrical utilities protection
   and cloud-based service. For each case, this document will identify
   the application, representative solutions used today and its
   requirements on deterministic latency mechanism.

Status of this Memo

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

   Internet-Drafts are working documents of the Internet Engineering
   Task Force (IETF), its areas, and its working groups. Note that
   other groups may also distribute working documents as Internet-Drafts.

   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."

   The list of current Internet-Drafts can be accessed at
   http://www.ietf.org/ietf/1id-abstracts.txt

   The list of Internet-Draft Shadow Directories can be accessed at
   http://www.ietf.org/shadow.html

   This Internet-Draft will expire on April 30, 2017.

Table of Contents

1. Introduction

Deterministic latency network (DLN) is defined to provide guaranteed
deterministic latency for specific traffic, especially for latency-
critical traffic. However, there are no current off-the-shelf
solutions for DLN. Distinguished from DetNet [I-D.finn-detnet-
architecture] that focuses on the solution of Layer 2-3 packet
forwarding, a feasible DLN solution is supposed to provide a latency
information obtaining mechanism including delay-aware modeling and
measurement, latency management mechanism with interfaces and
protocols to maintain the performance of latency sensitive
applications. Latency slicing and latency modeling are alternatives
in discussion currently. More specifically, DLN is:

-Face to upper layer, define delay-aware PHB and related OAM
interfaces;

-Targeting at WAN, support multiple data techniques including TSN;

-Working together with DetNet, and also have constraints on traffic flow, device capability, and etc.

This draft elaborates use cases from diverse industries along with their stringent requirements for deterministic low latency. Together, they provide broad industry context for DLN and a yardstick against which proposed DLN solutions can be measured.

For each use case, we identify the application with its latency requirement, and specify its representative solutions used today as well as problems to be settled on deterministic latency mechanisms.

1.1. 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].

2. Cloud VR/Gaming Service

2.1. Use Case Description

Virtual reality (VR) systems track one or more objects to generate the depiction of a virtual environment from the user's vantage point. Free viewpoint television (FTV) as a typical application of VR allows the user to interactively control the viewpoint and generate new views of a dynamic scene from any 3D position by transmitting all views of 3D space to user and rendering new images according user's turn locally, which requires large bandwidths, 1Gbps for 4K FTV and 4Gbps for 16K FTV, unavailable in subscriber network. VR gaming allows the user to experience being in a three-dimensional environment and interact with the environment during a game through dedicated equipments. Large bandwidth requirement and high cost of the customized hardware with immense computational power make VR inaccessible to the average consumer.

To address these challenges, cloud-based virtual reality is proposed by carriers to alleviate complex processing and high bandwidth on user side by outsourcing the computational power necessary to encode videos or render games to distant server farms and delivering only video content to user. Then carriers can serve as media providers in broadcast and streaming for sports and live events, or as

infrastructure provider for third-part VR applications. By reducing
initial cost of virtual reality and removing the need d for consumers
to constantly upgrade their equipments every few years, cloud based
virtual reality is much more attainable.

2.2. Delay in a VR system

In this use case we consider the latency between the time user start
to turn and the time the image is redrawn to account for the new pose
in a cloud based VR system. A high latency may induce simulator
sickness due to the virtual image drifting, reduce the subjective
sense of "presence", and change the pattern of behavior such that
users make more errors during speeded reaching, grasping, or object
tracking. To deliver good experiences, 20 ms of latency or much less
is recommended for a VR system to prevent simulator sickness.

The latency of a VR system mainly comes from the following steps:

1) Tracking has to capturing the exact pose of the HMD (Head Mount
Display), that is, the exact position and orientation in the real
world. Tracking latency is highly dependent on the system used. An
IMU (3-DOF gyro and 3-DOF accelerometer) has low latency on the order
of 1 ms, but drifts. Camera-based tracking doesn't drift, but has
high latency of 10-15 ms. Right now one of the lowest-latency non-
drifting accurate systems out there is a high-end system from NDI,
which has about 4 ms of latency. To reduce the tracking latency, the
obvious solution is using both optical tracking and an IMU, via
sensor fusion. The IMU can be used to provide very low-latency state,
and optical tracking can be used to correct the IMU's drift. Properly
implemented, sensor fusion can reduce the tracking latency to about
1.5 ms.

2) The graphic system has to render the scene as viewed from that
pose. Rendering latency depends on CPU and GPU capabilities and on
the graphics complexity of the scene being drawn. A VR system
requires at least 60 Hz for a good experience, and the corresponding
rendering latency is 16ms. When this step is moved to the cloud, the
rendering latency can be reduced to about 7.4 ms by distributed
parallel processing.

3) Transmission of the rendered output from the cloud to the user's
device over IP induces a transmit latency which is unpredicted
currently.

4) The graphics hardware has to transfer the rendered scene to the
HMD's display. This is called scan-out, and involves reading
sequentially through the frame buffer from top to bottom, moving

right to left within each scan line, and streaming the pixel data for
the scene over a link such as HDMI to the display. At 200Hz, the
displaying latency is on the order of 6ms.

The total latency of tracking, rendering and displaying is
1.5+7.4+6=14.9ms. So the suggested budget for transmit delay is 5ms
to keep the total latency down to 20 ms.

```
                                             +----------------------+
         +-----------------------------+     | Rendering Latency    |
         | +-------+     +--------+     |     |+-------+  +---------+|
         | |Sensors|-----|  A/D   |     |     ||cloud  |--|rendering||
         | +-------+     +---/----+     |  +----||process|  +---------+|
         |                   |          |  |   |+-------+      |     |
         | +---------+   +---/--------+| |   |+---------+ +--------+|
         | |terminal |---|signal      || |   ||video    |-|video   ||
         | |processing|   |transmission|| |   ||streaming| |encoding||
         | +----/-----+   +------------+| |   |+---------+ +--------+|
         | Tracking Latency            | |   +----------------------+
         +------|-----------------------+  +------------+
               |                           |network     |
               |                           |transmission|
            +--------------------------/             |
                                       |             |
                                       +------------+


                                       Transmit Latency

                                       +------------+
                                       |network     |
                                       |transmission|
                                       |            |
                                       +------------+
         +--------------------------+          |
         |   Displaying Latency     |          |
         | +----------+   +---/------+|         |
         | |          |---|Video     ||----------+
         | |displaying|   |decoding  ||
         | +----/-----+   +----------+|
         +--------------------------+
```
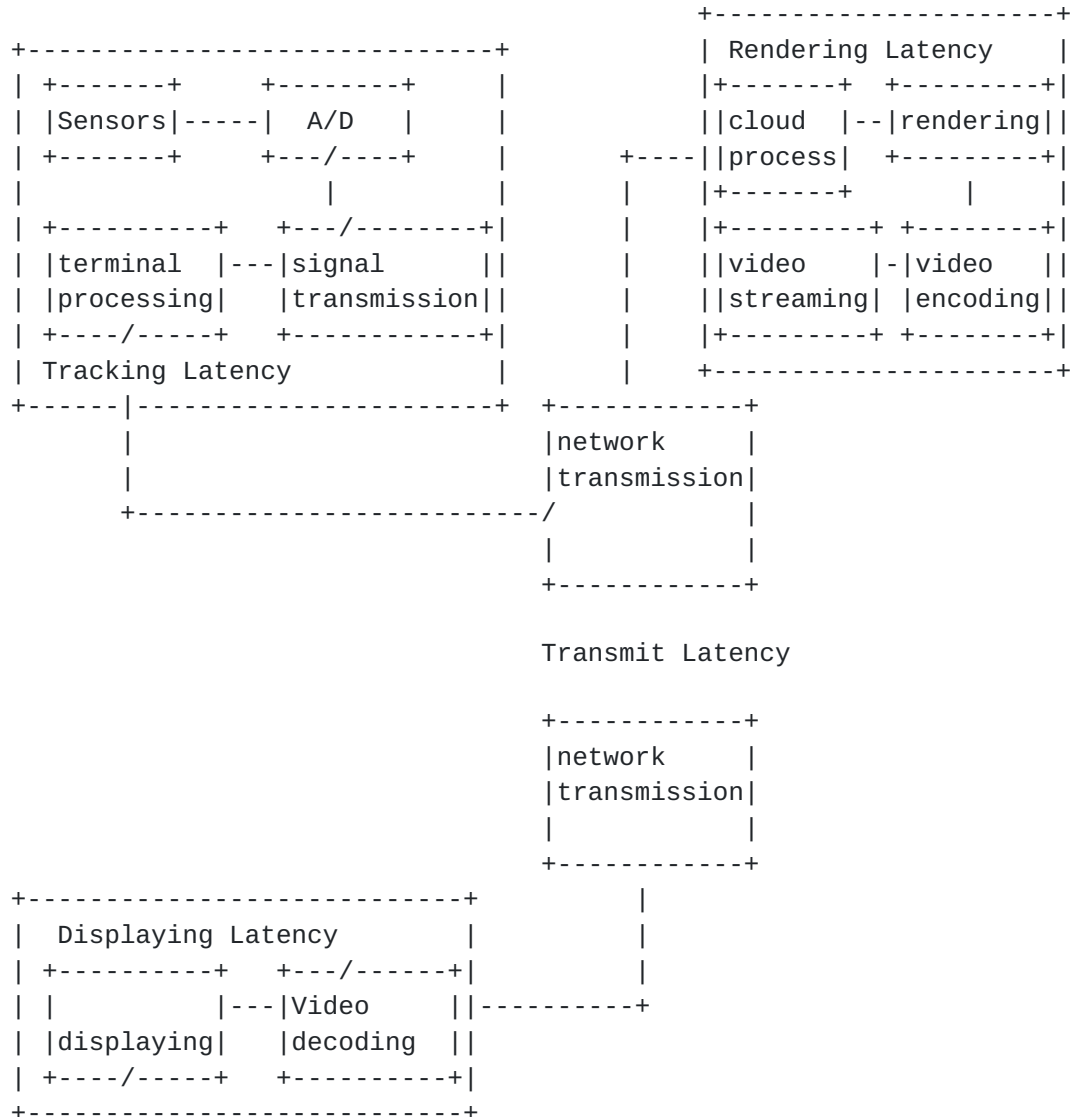
                        Figure 1 End-to-End Cloud VR System

Two deployments of cloud-based virtual reality are proposed in this
use case to keep the transmit latency down to 5ms. One is deploying
the central cloud access to the BRAS, and the other one is deploying
the central cloud access to the CO. There is a tradeoff between
latency and OPEX. Both of them require the network to guarantee a
deterministic low latency for VR traffic.

```
                                    /''''
                                    |     \
                                 /''''       \
                                /              \
                                |   VR Cloud   |
+-------+                        ///////////////
|   VR  |-----|                        |
|Devices|     |     +-----+    +-----+      +------+     +------------+
+-------+     |-----| ONT |----| OLT |------| BRAS |-----| Core Router|
              |     +-----+    +-----+      +------+     +------------+
+-------+     |
|   VR  |-----|
|Devices|
+-------+
```

                    Figure 2. VR Cloud on CO Position

```
                                    /''''
                                    |     \
                                 /''''       \
                                /              \
                                |   VR Cloud   |
+-------+                        ///////////////
|   VR  |-----|                        |
|Devices|     |     +-----+    +-----+      +------+     +-------------+
+-------+     |-----| ONT |----| OLT |------| BRAS |-----| Core Router |
              |     +-----+    +-----+      +------+     +-------------+
+-------+     |
|   VR  |-----|
|Devices|
+-------+
```
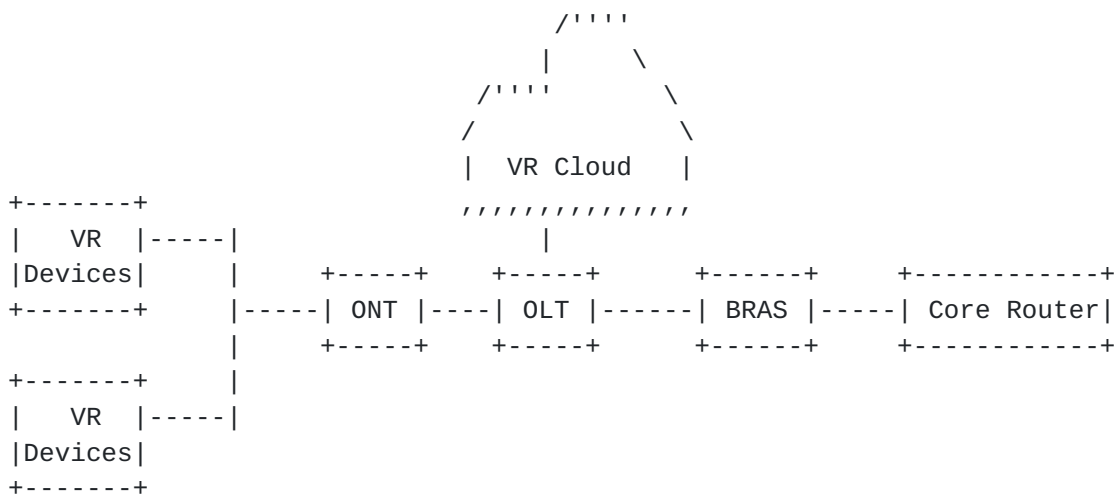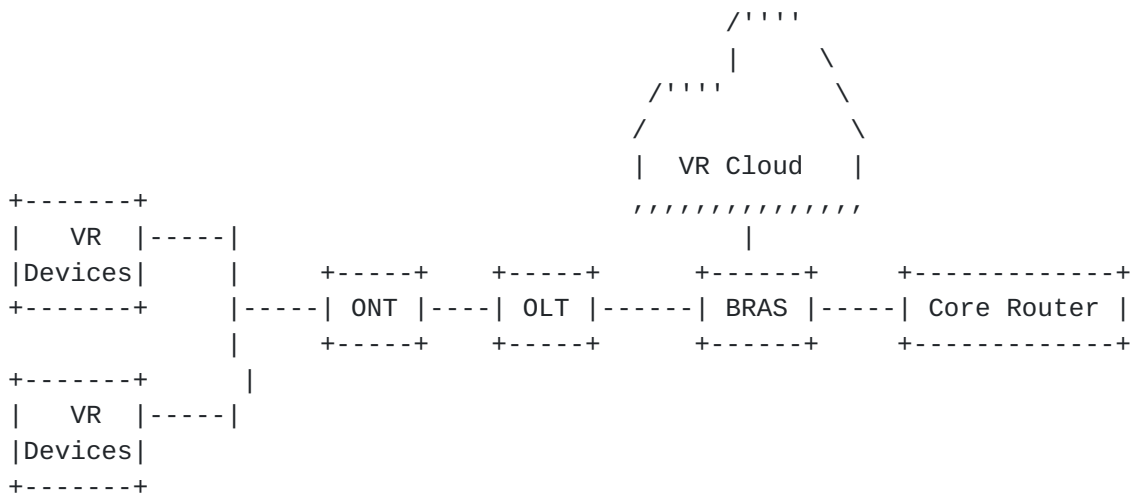
                    Figure 3. VR Cloud on Bras Position

In this case, assured transmission latency for VR traffic is the key
to cloud-based virtual reality. New latency-aware packet forwarding
mechanism is required for the network to guarantee the deterministic
latency for this latency-critical traffic. Along with this forwarding
mechanism, the latency is required to be measurable and manageable to
maintain the performance of VR.

2.3. Cloud VR/Gaming Asks

o Deterministic Delay behavior
o IP packet service
o Ultra low delay less than 5ms
o Delay management over network

3. Electrical Utility Relay Protection

3.1. Use Case Description

The effective operation of an electrical utility depends on high
validity and deterministic behavior of the fundamental networks.
Protection means not only to protect the operator but also to protect
the stability of the electrical equipment. If a transmission or
distribution power failure happens, it will cause damage to the
operator and large blackouts. The role of the protection system of
the electrical utility is to selectively trip out a faulty part by
delivering command signals as soon as possible.

The basic principal of protection scheme is that, by monitoring the
abnormal voltage or current changes in power primary equipment or
transmission lines, the logic program can identify a device or a
transmission line fault point and control circuit breaker to trip in
time. The current value is the same at both ends of the transmission
line during normal operation, and when this transmission line fails,
the current at both ends will not match. When the differential
current is greater than the setting value, the circuit breaker will
be tripped on each side of the protected device, so that faulty
equipment disconnects the power.AS is shown in Figure 1, A and B on
behaves of the line protection equipment, T1 and T2 refers to the
time delay of two opposite directions, Im and In refers to the current
of two opposite directions, Ik is the result of Im plus In, if Ik
equals to 0, there is no trouble between A and B, otherwise, some
faults are here.

```
        Im-------           Ik             -------In
         +---+   |           |           |   +---+
         | A |--|-----------|-------------|--| B |
         +---+   |           |           |   +---+
          |   |                            |   |
          |   |                            |   |
          |   +------------T1---------------+   |
         +-----------------T2------------------+
          /------------------------------------\
          |Normal or external trouble, Ik=Im+In=0|
          |internal trouble, Ik=Im+In 0          |
          \------------------------------------/
              /--------------------------\
              | T1 is the delay from A to B|
              | T2 is the delay from B to A|
              | Asymmetric=|T1-T2|         |
              \--------------------------/
           Figure 4 Electrical Utility Relay Protection
```
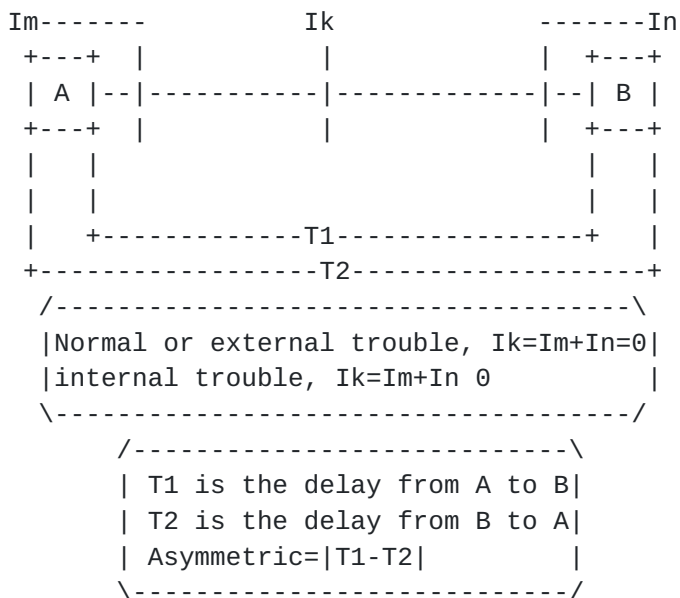
3.2. Delay Requirements on Protection Scheme

Right now everything is moving to IP network. In order to realize the
effective differential protection, time delay need to be taken into
consideration. As shown in Figure 1, T1 and T2 must be less than 5ms.
The most important thing is that the difference between T1 and T2
must be less than 20us. More detailed requirements are shown in the
following table:

```
+-----------------------------+-------------------------------+
|Power relay requirements      |Attributes                    |
+-----------------------------+-------------------------------+
|Interface type                |E1 interface                  |
+-----------------------------+-------------------------------+
|Full nodes                    |Less than or equal 15         |
+-----------------------------+-------------------------------+
|Total transmission distance   |About 500KM                   |
+-----------------------------+-------------------------------+
|One way delay                 |Less than 5ms                 |
|Maximum jitter                |10us                          |
+-----------------------------+-------------------------------+
|Asymmetric delay              |Less than 20us                |
+-----------------------------+-------------------------------+
```

Table 1: Power Relay Requirements

3.3. Precision Delay Compensate Technology

The effective differential protection relies on synchronous sampling.
It means that the protective device on both sides of the transmission
line requires synchronous sampling. When both sides of relay
protection exists sampling time difference, it will produce a
corresponding differential current value. If the differential current
value exceeds the threshold value, the protection may miss operate.
Synchronous sampling is based on the consistency of the two-way
channel delay, which implies that the smaller the asymmetric is, the
better differential protection behaves. So the relay protection has
strict requirements on asymmetric delay, it must be less than 20us.

SDH technology has become the mainstream of the electric power
communication network technology. In recent years, with the process
of smart grid construction accelerated, more variety, a greater flow
of business types gradually add in electric power communication
network. It becomes more and more difficult for SDH network to
satisfy time delay in the differential protection. So the main
electric power communication network service is from the traditional
TDM business gradually transformed into IP packet service. The highly
precise time delay compensate technology plays an important role in
IP packet service, which can solve the problem of time delay in the
differential protection.

3.4. Relay Protection Asks

o Deterministic behavior
o Synchronous sampling
o IP packet service
o 20us Asymmetric delay
o Delay information for compensation technology

4. Cloud-based Service

4.1. Use Case Description

Cloud-based services those are any resources provided over the
network. Commonly, the cloud providers provide computing, storage as
well infrastructures as a service which is accessible from network.
More and more companies are switching to the cloud since it is more
efficient, secure and reliable. The performance is the key matter of
cloud-based services, as fast and predictable response from remote
location at any volume is the determinant of user experience.

Latency and loss is the killer of cloud-based services with a direct
negative impact on performance. There are many factors can affect
latency, distance, routing, bandwidth and so on. CDN and deployment
of edge caches can help to mitigate some of the delays but only for
downloads and do little for upstream traffic. Delay and jitter on
video conferences or poor application performance still remains on
where traffic goes.

In this case, all cloud providers guarantee a minimum service level
to users. Users will be refunded in certain level if there is a
violation on SLA. The SLA is usually based on availability
requirements on a monthly or yearly basis from 99% to 100%. From [],
some operators guarantee uptime, but current no SLA is based on
performance indicators such as response time. The reason below this
is lack of such delay control and management mechanisms.

Visible and aware of delay performance is also necessary to maintain
a global view of state of art of the network.

4.2. Cloud-based Service Asks

o Deterministic Delay behavior
o IP packet service
o Delay visibility
o Delay management over network

5. Security Considerations

This document analyzes the standardization work on synchronization in
different SDOs. As no solution is proposed in this document, no
security concerns are raised.

6. IANA Considerations

There are no IANA actions required by this document.

7. References

7.1. Informative References

[I-D.finn-detnet-architecture]

Finn, N., Thubert, P., and M. Teener, "Deterministic Networking
Architecture", draft-ietf-detnet-architecture-00 (work in progress),
September 2016.

8. Acknowledgments

   TBD

Authors' Addresses

   Xian Liu
   Huawei Technologies Co., Ltd.
   Bantian, Longgang district
   Shenzhen 518129, China
   Email: lene.liuxian@huawei.com