Use cases of Application-aware Networking (APN) in Edge Computing
draft-liu-apn-edge-usecase-03

Abstract

   The ever-emerging new services are imposing more and more highly
   demanding requirements on the network. However, the current
   deployments could not fully accommodate those requirements due to
   limited capabilities. For example, it is difficult to utilize the
   traditional centralized deployment mode to meet the low-latency
   demand of some latency-sensitive applications. Moreover, the total
   amount of centralized service data is growing exponentially, which
   brings great pressure on the network bandwidth. There has been a
   clear trend that decentralized sites comprising of computing and
   storage resources are deployed at various locations to provide
   services. In particular, when the sites are deployed at the network
   edge, i.e. the Edge Computing, it can better handle the business
   needs of the users nearby, which provides the possibilities to
   provide differentiated network and computing services. In order to
   achieve the full benefits of the edge computing, it actually implies
   a precondition that the network should be aware of the applications'
   requirements in order to steer their traffic to the network paths
   that can satisfy their requirements. Application-aware networking
   (APN) aims to accommodate the edge services' needs, fully releasing
   the benefits of the edge computing.

   This document describes the various application scenarios in edge
   computing to which the APN can be beneficial, including augmented
   reality, cloud gaming and remote control, which empowers the video
   business, users interaction business and user-device interaction
   business. In those scenarios, APN can identify the specific
   requirements of edge computing applications on the network, process
   close to the users, provide SLA guaranteed network services such as
   low latency and high reliability.

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].

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 https://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 4, 2021.

Copyright Notice

Table of Contents

1. Introduction

   Edge computing is to deploy service sites near the user side to
   provide users with better network and computing services. The
   services of edge computing can not only be implemented in the edge
   data center, but also be integrated in the network equipment, which
   brings the possibility for the convergence of network and computing,
   and also puts forward the requirements for the technology combining
   of different industries. On the one hand, the demand of different
   applications for the network need to be exposed; on the other hand,
   the network needs to be aware of computing power and steers the
   traffic along the appropriate path towards the suitable sites.

   The existing network can only identify the application demands in a
   coarse granularity. When the application demand is high causing the
   heavy network load, it usually fails to guarantee the latency and
   reliability of the applications especially the mission-critical
   applications. Application-aware networking (APN) faciliates service
   provisioning in a fine granularity, and then either steer the
   corresponding traffic onto the appropriate network path (if exist)
   that can satisfy these requirements or establish an exclusive network
   path which wouldn't be influenced by other applications' traffic
   flow.

2. Edge Computing and APN

   In a whole edge computing network, there are user terminal, edge
   gateway and edge data center. The edge gateway can be the UPF In 5G
   network. Edge data center is usually close to users and serves a
   limited group of users, the network and computing tasks performed by
   edge computing are more specific and customized. Both computing
   resources and network resources need to be able to provide fine-
   grained service guarantee. The goal of APN is to provide fine-

grained network service, including latency, jitter, reliability and others, which can be well matched with edge computing.

Appilication-aware networking includes the app-aware edge (APN-Edge), app-aware process head-end (APN-Head), app-aware process mid-point (APN-Midpoint) and app-aware process end-point (APN-Endpoint). A user's request is sent from the client, and then passes through all the nodes of the APN network to the server. The function of APN-Edge can be deployed in the edge gateway, so the request traffic of client can be distinguished by the edge gateway/APN-Edge and sent to the edge data center through the APN. In some cases, the reply of the edge data center will not return to the original client, and may be sent to another client through the APN.

```
    +------+    +----------------+    +-------------+    +---------+
    |      |    | Edge Gateway/  |    |     APN     |    | Edge    |
    |Client|<-->|                |<-->|             |<-->| Data    |
    |      |    | APN-Edge       |    |   Network   |    | Center  |
    +------+    +----------------+    +-------------+    +---------+
```
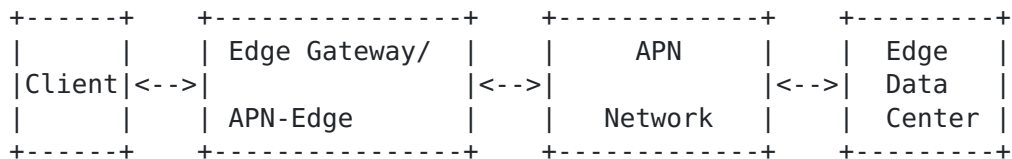
                    Edge Computing and APN

3. Usage Scenarios of APN in edge computing

This section presents several typical scenarios which require edge computing to interconnect and to co-ordinate with APN to meet the service requirements and ensure user experience.

3.1. Augmented Reality (AR)

3.1.1. Use Case Description

Augmented reality is a relatively new application that promotes the integration of real world information and virtual world information content. It includes several technologies, such as track registration, display, virtual object generation, interaction and merging.

3.1.2. Augmented Reality Today

AR gives users an immersive experience. It is widely used in the consumer industry presently, and may also be applied in industrial fields such as health care and education in the future. The general process of AR / VR is as follows:

* Image acquisition equipment (such as camera) collects image or video information and sends it to data center.

* Data center carries out identification, feature extraction and
template rendering, and sends them to AR terminal.

* The AR terminal plays the synthesized information.

Considering the user experience, AR usually needs a high bandwidth of
100mbps due to multi-channel acquisition of image or video data, and
a low end-to-end latency less than 60ms. With centralized
deployment, the network transmission distance is too long, so the
latency demand can't be met; the large volume of traffic load also
imposes high challenge on the network bandwidth.

3.1.3. Augmented Reality with Edge Computing and APN

If the deployment mode of edge computing is adopted, the following
functions can be realized:

* The collected image or video information can be encoded/decode and
compressed by the edge equipment to reduce the bandwidth requirements
of data transmission.

* The edge data center can process the collected image or video data
nearby and send it to the AR terminal equipment, which reduces the
distance of network transmission and greatly reduces the latency.

Although edge computing can reduce the overall latency of services
and reduce the demand for network bandwidth, it still needs
differentiated network services to provide the ultimate guarantee for
application with high SLA requirements. APN can achieve:

* Edge device obtains and encapsulates AR application feature
information and sends it to the headend node.

* Headend node in the APN identifies the AR data flow and steers it
into a specific transmission path according to the demanded
bandwidth, latency and reliability.

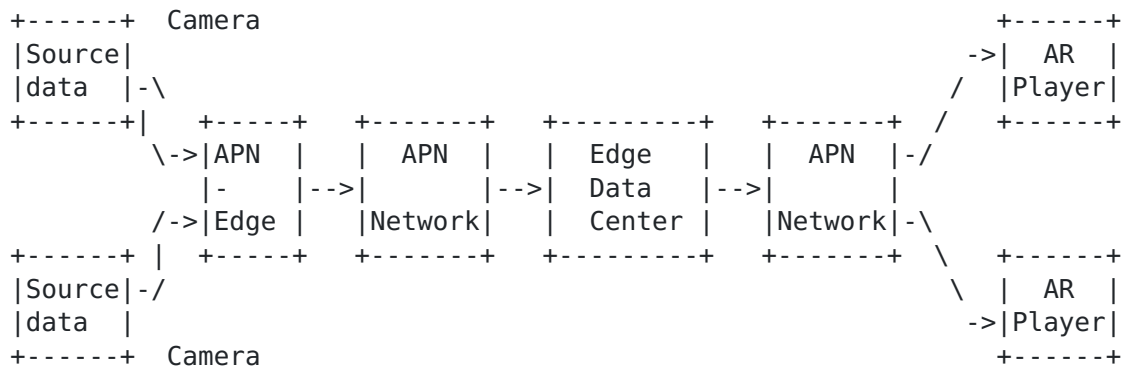* Mid point in the APN forwards the data stream along the specific
path.

* End point in the APN receives AR data stream and forwards it either
to Data Centre for processing or to the AR player for playing.

In the whole process, because APN identifies the traffic of AR
application, it can provide corresponding network services to provide
customized high reliability, low latency and other SLA guarantee.

```
+------+  Camera                                        +------+
|Source|                                            ->|  AR  |
|data  |-\                                         /  |Player|
+------+|   +-----+   +-------+   +---------+   +-------+  /   +------+
        \->|APN  |   |  APN  |   |  Edge   |   |  APN  |-/
           |-    |-->|       |-->|  Data   |-->|       |
        /->|Edge |   |Network|   | Center  |   |Network|-\
+------+ | +-----+   +-------+   +---------+   +-------+  \   +------+
|Source|-/                                           \  |  AR  |
|data  |                                           ->|Player|
+------+  Camera                                        +------+
```

                Augmented Reality with Edge Computing and APN

3.2. Cloud Gaming

3.2.1. Use Case Description

   Cloud gaming is to deploy the game application in the data center,
   and realize the functions includes the logical process of game
   command control, as well as the tasks of game acceleration, video
   rendering and other tasks with high requirements for chips. In this
   way, the terminal is a video player. Users can get a good game
   experience without the support of high-end system and chips.

   Compared with the traditional game mode, there are several advantages
   of cloud game, such as no installation, no upgrade, no repair, quick
   to play and reduce the terminal cost, so it will have stronger
   promotion.

3.2.2. Cloud Gaming Today

   The biggest feature of cloud games is that users interact with each
   other through the network. The general process is as follows:

   * The data center sends game video streaming information to the
   terminal, including game background picture, characters, etc.

   * The user makes corresponding operation instructions according to
   the received game video stream information and sends them to the data
   center.

   * The data center constantly updates the video stream and other data
   of the game according to the user's operation instructions.

   Game users usually pursue consumption experience. Currently, most
   users are willing to spend extra money in order to obtain better user
   experience. Generally speaking, the network latency of game is

required to be less than 30ms. For competitive game, the latency
will be required to be less than 10ms, because professional players
usually can feel the millisecond level latency difference. With
centralized deployment, the network transmission distance is too
long, which is a huge challenge to the network load, so the latency
demand can't be met; the large volume of traffic load also imposes
high challenge on the network bandwidth.

3.2.3. Cloud Gaming with Edge Computing and APN

If the deployment of edge computing is adopted, the following
functions can be realized with the deployment of edge data center:

* The edge data center sends the game video stream information to the
terminal, and receives the user's control instruction information for
processing.

* users can make corresponding operation instructions according to
the received video stream information, and get quick response.

Edge computing can reduce the latency of game data transmission as a
whole, but it should be noted that cloud games usually have multiple
players playing a game together, which requires the deterministic
latency of multi-party network path, which needs to be realized with
APN:

* Multiple edge devices obtain and encapsulate cloud game application
feature information and send it to the head end node.

* Headend node in the APN identifies the data flow of cloud games
(maybe the same game), and steers it into a specific transmission
path according to its requirements for bandwidth, delay, reliability,
etc., which needs to ensure that the latency of multi-user control
instructions arriving at the edge data center is consistent.

* Midpoint in the APN forwards game data stream according to the
predetermined path.

* The endpoint in the APN receives the cloud game data stream and
steers it either to the data center for processing the users' control
instruction or to the user for playing.

The whole process requires APN not only to identify the cloud game
traffic and provide customized network forwarding services for it,
but also to ensure the deterministic latency of multi-user in the
same game and provide better game experience.

```
 Client A
+---------+
|Game data|
+---------+-\   +----------+   +-----------+   +-----------+
           |<->|   APN-    |-A-|    APN     |-A-|           |
           |   | Edge A   |   | Network A |   |           |
           +----------+   +-----------+   | Edge Data |
           +----------+   +-----------+   |  Center   |
           |   APN-    |   |    APN     |   |           |
           |<->|  Edge B  |-B-| Network B |-B-|           |
+---------+-/   +----------+   +-----------+   +-----------+
|Game data|
+---------+
 Client B
```
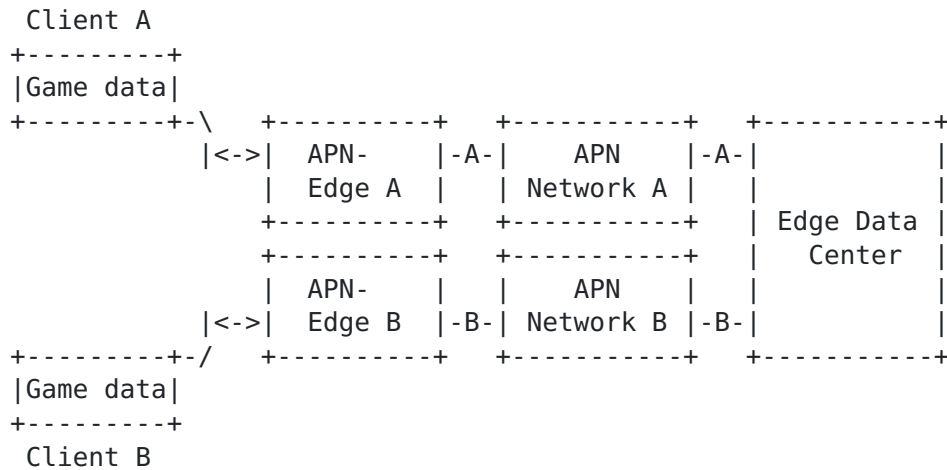
               Cloud Gaming with Edge Computing and APN

3.3. Remote control of industry

3.3.1. Use Case Description

Industrial remote control refers to the remote control of field
equipment in areas that are not convenient for manual field control,
such as high-temperature and high-risk areas. In the past, signaling
was usually transmitted through industrial private networks and
protocols. With the development of industrial Internet, the industry
also gradually has the demand of network interconnection. Its
network tends to adopt L3 protocol and flat architecture, which makes
it possible for cross distance remote control service.

3.3.2. Remote control of industry Today

In the process of remote control, workers constantly make control
instructions according to the received image or video information of
field equipment, which requires interaction between personnel and
equipment through the network. Because the field environment that
needs remote control is generally poor, it is also a challenge for
the security of the operation equipment. If the latency is too large
or the reliability is not enough, it may cause the operation failure,
equipment damage and other serious consequences. Therefore, the
remote control service requires low latency and high reliability.
The general process of remote control is as follows:

* Field equipment (such as camera) collects image or video
information and sends it to data center.

* The data center receives the field information of the equipment and
sends it to the workers in the office.

    * Workers send control instructions and control equipment according
    to the received field information.

    Many industrial enterprises rent public cloud resources to construct
    their own data center, but the long distance of network transmission
    is not conducive to the timely transmission of image / video data
    stream, which will cause large latency and packet loss.

[3.3.3](#). Remote control of industry with Edge Computing and APN

    If the deployment mode of edge computing is adopted, and the data
    center and edge computing access equipment (such as gateway) are
    deployed in a location or enterprise park close to the business site,
    the following functions can be realized:

    * The collected image or video information can be encoded/ decoded
    and compressed by edge access equipment to reduce the bandwidth
    requirements.

    * The control instruction information can be identified by the edge
    equipment, so as to provide exclusive network transmission service.

    * The forwarding path of image / video and control information is
    shortened, which can greatly reduce the latency.

    Although edge computing can reduce the overall delay of services and
    reduce the demand of network bandwidth, it still needs to achieve
    differentiated network services through APN to provide the ultimate
    network guarantee for the services with the highest network
    requirements.

    For users, APN can realize those functions.

    * Edge device obtains and encapsulates the image or video information
    of the remote field device, then sends it to the headend node.

    * Headend in the APN identifies the information and steers the flow
    into a specific transmission path according to its requirements for
    bandwidth, delay, reliability, etc..

    * Midpoint in the APN forwards along the specific path.

    * Endpoint receives image or video data stream of field equipment and
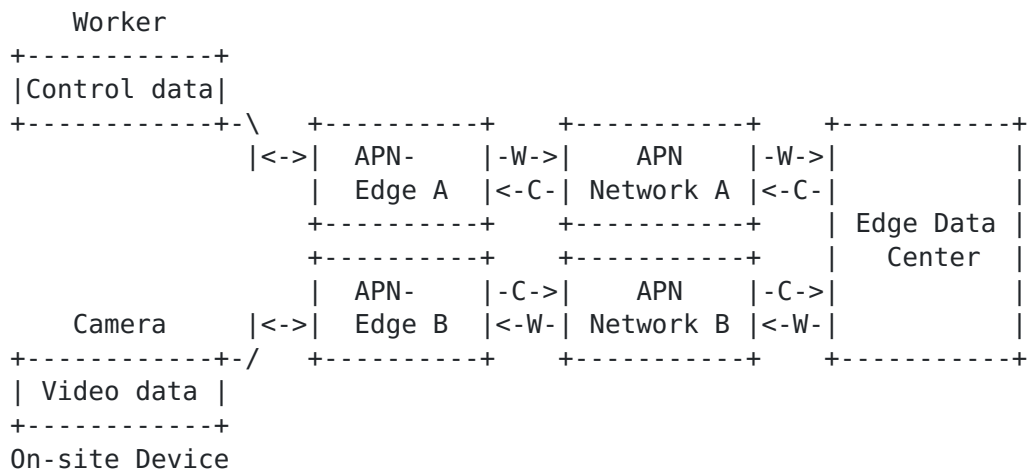   forwards it to users.

    For field equipment, APN can realize those functions.

* Edge device obtains and encapsulates the control instruction
  information and sends it to the head end node.

* Headend in the APN identifies the control data flow and steers into
  a specific transmission path according to the demand for bandwidth,
  latency and reliability.

* Midpoint in the APN forwards along the specific path.

* Endpoint receives control information and forwards to the field
  equipment.

In the whole process, APN identifies the traffic of remote control
service, which can provide customized high reliability, low latency
and other network guarantee.

```
      Worker
  +------------+
  |Control data|
  +------------+-\   +----------+    +-----------+    +-----------+
                |<->|  APN-    |-W->|    APN    |-W->|           |
                |   | Edge A   |<-C-| Network A |<-C-|           |
                +----------+    +-----------+    | Edge Data |
                +----------+    +-----------+    |   Center  |
                |  APN-    |-C->|    APN    |-C->|           |
     Camera     |<->| Edge B   |<-W-| Network B |<-W-|           |
  +------------+-/   +----------+    +-----------+    +-----------+
  | Video data |
  +------------+
  On-site Device
```

         Remote control of industry with Edge Computing and APN

[4]. Conclusion

APN enables low latency and high reliability network services in
various edge computing scenarios such as AR, cloud gaming, remote
industrial control, etc.

[5]. Security Considerations

TBD.

[6]. IANA Considerations

TBD.

[7](). Normative References

[I-D.li-apn-framework]
          Li, Z., Peng, S., Voyer, D., Li, C., Liu, P., Cao, C.,
          Ebisawa, K., Previdi, S., and J. N. Guichard,
          "Application-aware Networking (APN) Framework", [draft-li-apn-framework-02]() (work in progress), February 2021.

[I-D.li-apn-problem-statement-usecases]
          Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z.,
          Ebisawa, K., Previdi, S., and J. N. Guichard, "Problem
          Statement and Use Cases of Application-aware Networking
          (APN)", [draft-li-apn-problem-statement-usecases-01]() (work
          in progress), September 2020.

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

Authors' Addresses

   Peng Liu
   China Mobile
   Beijing 100053
   China

   Email: liupengyjy@chinamobile.com

   Zongpeng Du
   China Mobile
   Beijing 100053
   China

   Email: duzongpeng@chinamobile.com

   Shuping Peng
   Huawei
   Beijing 100053
   China

   Email: pengshuping@huawei.com

   Zhenbin Li
   Huawei
   Beijing 100053
   China

   Email: lizhenbin@huawei.com