Network Working Group
Internet Draft
Intended status: Standards Track
Expires: April 10, 2026

Y. Liu
R. Pang
China Unicom
October 11, 2025

Use Cases and Requirements of Massive Data Transmission（MDT） in High Bandwidth-delay Product (BDP) Network

draft-liu-rtgwg-mdt-in-high-bdp-02

Abstract

This document describes the use cases and related requirements of Massive Data Transmission（MDT） in High Bandwidth-delay Product (BDP) Network. The MDT framework enables efficient use of nighttime idle bandwidth to provide services such as same-day or next-day delivery. To meet these objectives, the system introduces a terminal-driven intelligent parameter optimization method that adjusts local configurations (e.g., disk I/O, NIC bandwidth, TCP buffer) to match transmission goals. This document outlines the end-to-end architecture and key mechanisms including service identification and traffic statistics, network layer load balancing, transmission protocol optimization, collaboration between terminal APP, network device and controller, etc.

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), 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 10, 2026.

Table of Contents

1. Introduction

With the continuous development of industries such as autonomous driving, AI intelligent computing, and enterprise cloud, the demand for massive data transmission across wide area networks from edge data centers/enterprises to core data centers has become increasingly common, and higher requirements have been put forward for existing carrier network architectures.In particular, the implementation of the national "East-to-West Computing Resource

Transfer" strategy has significantly increased the frequency and scale of long-distance data migration. Typical scenarios involve the transfer of massive scientific computing data—such as irregularly generated film production materials, periodically uploaded enterprise backup data, and daily astronomical data from facilities like FAST—over thousands of kilometers to intelligent computing centers located in western regions. These data flows are often insensitive to real-time latency but are highly sensitive to overall delivery deadlines.

Taking the scenarios of supercomputing and intelligent computing as example, data transmission usually includes two requirements:

1) The transmission of training data between intelligent computing centers, supercomputing centers, and between intelligent computing centers and supercomputing centers is usually carried by optical networks due to high bandwidth requirements and high connection stability.

2) The transmission of training data and result feedback between users and intelligent computing centers/supercomputing centers can be carried through IP networks due to their strong suddenness and cost sensitivity.

While traditional private leased lines can support such Massive Data Transmission (MDT) demands with reliable bandwidth guarantees, they often involve rigid billing (e.g., daily/monthly rates), high costs, and underutilized fixed capacity. As a result, they fail to meet the flexibility requirements for transmitting large-scale cold or semi-cold data that prioritize cost efficiency and deadline-based delivery. This leads many data providers to resort to physical shipment of hard disks as a practical alternative.

To address these limitations, MDT is proposed as a flexible, time-sensitive service model that utilizes idle network bandwidth (such as during nighttime) to deliver large-scale data efficiently—enabling offerings like same-day or next-day delivery guarantees. Furthermore, terminal-side configuration (including disk type, network interface card, and TCP buffer size) is found to significantly affect transmission performance in high Bandwidth-Delay Product (BDP) networks.

This draft mainly describes the overall architecture of feasible solutions for MDT in high BDP network, typical problems that may be encountered, and proposes potential solutions, including but not limited to how to perform load balancing scheduling at the global level of the network to avoid the impact of massive data transmission on existing network services; how to identify MDT

services for traffic record and billing purposes; how to optimize the congestion control algorithm of the transport layer protocol to ensure that the throughput of TCP protocol can be improved in long-distance lossy networks.Based on this, this draft introduces an architecture and supporting mechanisms for MDT, including: a terminal-driven intelligent parameter tuning method that dynamically adjusts local configurations to meet transmission goals; adaptive transmission protocol optimizations (e.g., congestion control tuning); collaborative scheduling and decision-making mechanisms across endpoints, network controllers, and transport devices.

1.1. 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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Massive data transmission in high BDP network

Figure 1 show schematic diagram of the network architecture of MDT in high BDP network, where key functional units include:

DC/User APP：APP can be deployed on DC/personal terminal devices, which can be traditional file transfer tools or customized apps developed for MDT scenarios, which implements enhanced functions such as intelligent data compression, intelligent partitioning, encryption, etc.

Network：existing service carrier networks of current operators, such as metropolitan area networks, backbone networks, etc.

Controller：existing network management system includes controllers, collaborators, orchestrators, etc, responsible for real-time resource scheduling, configuration orchestration, and feedback loops.

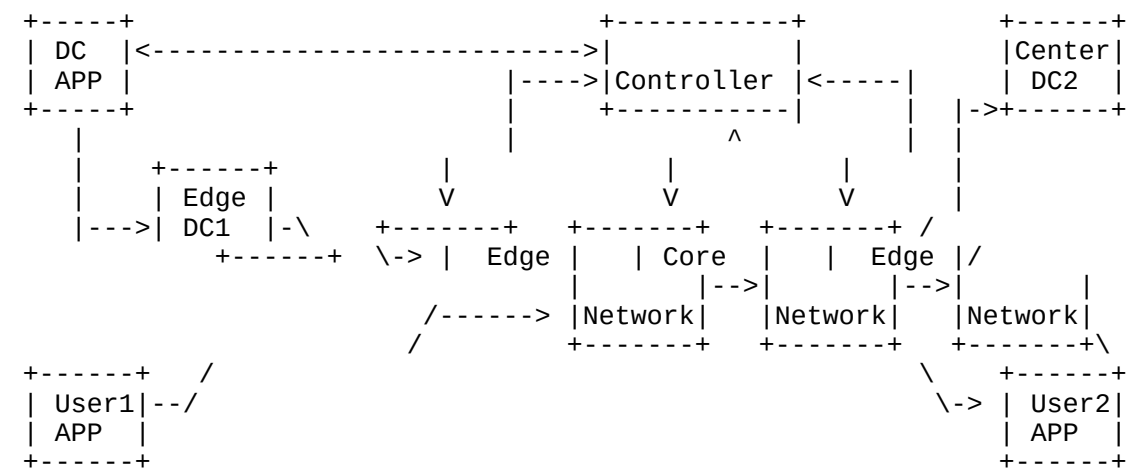

Figure 1: Architecture of MDT in high BDP network

3. Use Cases and Requirements

MDT service is a predictable time-efficient service that requires data transmission to be completed within a specified time, not sensitive to transmission delay, but requires a considerable amount of network resources. Compared with traditional Internet and private line services, how to improve transmission efficiency, achieve service identification, complete scheduling and billing for services are key issues to be considered.

3.1. Service Identification and Traffic Record

Before starting transmission, the terminal client collects environmental parameters (disk speed, NIC bandwidth, TCP buffer size) and interacts with the controller to communicate task objectives (data volume, completion time).The controller will dynamically adjust the network path calculation and private line bandwidth based on transmission requirements and the current available link resources of the network, and distribute the configuration to network nodes.

After the transmission task is initiated, network devices need to be able to identify MDT services and corresponding account information based on certain identifiers, perform traffic record, and report the statistical results to the controller. The controller can get the overall MDT usage of the network, as well as the specific resource completion status for a particular user, and make corresponding strategy adjustments. This feedback enables real-time adjustment and facilitates multi-path selection based on residual bandwidth rather than only latency.

From the above use cases, it can be seen that the billing of MDT services and the scheduling and allocation of available bandwidth resources in the current network require network devices to recognize the current MDT services. APNID defined in [I-D.li-apn-problem-statement-usecases] may be a potential solution to meet the identification requirements of MDT service.

3.2. Load Balancing at Network Level

The bandwidth requirement for MDT service is generally between 500M-10G, and the launch of each service requires a huge consumption of network resources. With the continuous increase of service launching, how to make reasonable use of network idle resources, allocate global network resources and data express tasks, and minimize the impact on existing services have become necessary issues to consider.

The controller notifies the APP of available network resource information, and the APP dynamically adjusts the data sending strategy based on the available network bandwidth, and cooperates with network devices to improve the overall resource utilization of the network. When the controller discovers a shortage of available network resources or predicts a rapid growth in future network traffic, it should notify the APP side in advance to make policy adjustments.

Instead of defaulting to shortest-latency paths, MDT optimizes for highest bottleneck bandwidth. For example, if two routes exist between Shanghai and Ningxia—one with lower latency but 1Gbps bottleneck, and another with 5Gbps bottleneck—the MDT system prefers the latter to minimize transmission time. This route selection strategy is influenced by application-level QoS constraints, such as delivery deadlines, which are translated into required average throughput. The controller provides resource-aware path evaluations, and the terminal selects paths that satisfy task goals under cost constraints.

This approach allows MDT services to maximize bandwidth utilization during off-peak hours, reduce transmission costs, and avoid contention with latency-sensitive services such as video streaming or real-time communications.

3.3. Optimization of Transmission Protocols

In most scenarios, the two ends of MDT services need to cross a wide area network, with a distance of over 1000 km. RTT is in the tens of MS range, and there is a small amount of packet loss in the network, which poses new challenges to the traditional TCP [RFC7805]. Based on current test results, the traditional TCP congestion control algorithm [RFC2581] may not achieve the expected transmission rate for MDT. Therefore, an efficient, secure transmission protocol that can adapt to the current network state and resource status is needed to solve these problems. The current network satae, including bandwidth, congestion state, path, etc, can be measured by ippm and other network state measurement

technologies. The network state can be carried by the ack of transmission protocols (TCP or QUIC) and also processed in reciever. By using the network state information, the transmission protocols can achieve congestion control and traffic control to optimize the network throughput.
A key innovation is dynamic adjustment of the TCP buffer size using an iterative algorithm. The system compares the actual average throughput Ro with the target throughput, and iteratively adjusts buffer size until the transmission completion time meets the requirement. Cost estimation is also incorporated, using the formula: C = β t + τ B / 50MB, where β is the unit time cost, and τ is the per-50MB bandwidth cost. These parameters are selected from environment configurations such as network interface speed, disk read/write rate, and measured RTT, and refined through historical samples and machine learning.

The adjustment process supports both static initialization at the start of the task and dynamic updates during transmission. This ensures that TCP/QUIC can reach near-optimal throughput even in high BDP networks with limited user configuration effort.

3.4. Collaboration Requirements

To meet the resource requirements of MDT business while not affecting existing network services, efficient collaboration between controllers, terminal APPs, and network devices is required.

3.4.1. Collaboration Between Terminal APP and Controller

The terminal APP needs to authenticate and authorize the controller to accept scheduling. The APP reports the request for transmission tasks to the controller, such as the size of the data to be transmitted, the transmission completion time, the service priority, etc. The controller recommends one or more configurations, each including disk/NIC combo, TCP buffer size, start time, estimated time, and cost. Users may accept risk reminders and receive alerts during execution to switch plans when necessary.

3.4.2. Collaboration Between Network Device and Controller

The controller needs to get the real-time status of the network devices and the available bandwidth resources of all links, perform global routing and bandwidth configuration changes based on the service demands reported by the APP, and distribute the configuration to network devices. Network devices report their operational status, available resource information, and specific MDT

service information to the controller, and receive policy information provided by the controller.

3.4.3. Collaboration Between Network Device and Terminal APP

The host-network coordination scheme has been increasingly widely discussed. In this scheme, the network side needs to use some status information of network nodes, such as CIR, MTU, Link Usage (defined in RFC 9473), status information of Segment list, etc., and interprets traffic class markers to prioritize MDT vs. background flows,to guide the host(APP) to adjust the sending strategy.

The overall principle is that in the scenario of massive data transmission, the host can select the idle network paths and the most economical transmission mode flexibly based on the information provided by the network side, then improve the load utilization of the network.

At the same time, the host needs to notify the network side through some identification to indicate the APP status , and the key demands for the network node. The network devices can adjust the strategy based on this information. For example, in order to reduce the impact on the existing services of the current network, the priority of some less important service messages should be marked as not very important. Thus, when congestion occurs, the network equipment will preferentially discard these messages.

3.4.4. Interface Requirements

TBD.

4. Security Considerations

TBD.

5. IANA Considerations

TBD.

6. References

6.1. Normative References

[RFC7805] Zimmermann, A., Eddy, W., and L. Eggert, "Moving Outdated TCP Extensions and TCP-Related Documents to Historic or Informational Status", RFC 7805, DOI 10.17487/RFC7805, April 2016, <https://www.rfc-editor.org/info/rfc7805>.

[RFC2581] Allman, M., Paxson, V., and W. Stevens, "TCP Congestion Control", RFC 2581, DOI 10.17487/RFC2581, April 1999, <https://www.rfc-editor.org/info/rfc2581>.

[RFC 9526] Filsfils, C., Talaulikar, K., Ed., Voyer, D., Bogdanov, A., and P. Mattes, "Segment Routing Policy Architecture", RFC 9256, DOI 10.17487/RFC9256, July 2022, <https://www.rfc-editor.org/info/rfc9256>.

6.2. Informational References

[I-D.li-apn-problem-statement-usecases] Li, Z., Peng, S., Voyer, D., Xie, C., Liu, P., Qin, Z., and G. S. Mishra, "Problem Statement and Use Cases of Application-aware Networking (APN)", Work in Progress, Internet-Draft, draft-li-apn-problem-statement-usecases-08, 3 April 2023, <https://datatracker.ietf.org/doc/html/draft-li-apn-problem-statement-usecases-08>.

[I-D.li-apn-header] Li, Z., Peng, S., and S. Zhang, "Application-aware Networking (APN) Header", Work in Progress, Internet-Draft, draft-li-apn-header-04, 12 April 2023, <https://datatracker.ietf.org/doc/html/draft-li-apn-header-04>.

Authors' Addresses

Ying Liu
China Unicom
China
Email: liuy619@chinaunicom.cn

Mengyao Han
China Unicom
China
Email: hanmy12@chinaunicom.cn

Zheng Ruan
China Unicom
China
Email: ruanz6@chinaunicom.cn