

Internet Engineering Task Force (IETF)
Request for Comments: 9808
Category: Standards Track
ISSN: 2070-1721

A. Ryan
Disney Streaming
B. Rosenblum
Vecima
N. Sopher
Qwilt
July 2025

Content Delivery Network Interconnection (CDNI) Capacity Capability Advertisement Extensions

Abstract

This specification defines a set of additional Capability Objects that provide information about current downstream CDN (dCDN) utilization and specified usage limits to the delegating upstream CDN (uCDN) in order to inform traffic delegation decisions.

This document supplements the CDNI Capability Objects, defined in RFC 8008 as part of the Footprint & Capabilities Advertisement Interface (FCI), with two additional Capability Objects: FCI.CapacityLimits and FCI.Telemetry.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <https://www.rfc-editor.org/info/rfc9808>.

Copyright Notice

Copyright (c) 2025 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Revised BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Revised BSD License.

Table of Contents

1. Introduction
 - 1.1. Terminology
 - 1.2. Requirements Language
 - 1.3. Objectives
2. CDNI Additional Capability Objects
 - 2.1. Telemetry Capability Object
 - 2.1.1. Telemetry Source Object

2.1.1.1.	Telemetry Source Types
2.1.1.2.	Telemetry Source Metric Object
2.1.2.	Telemetry Capability Object Serialization
2.2.	CapacityLimits Capability Object
2.2.1.	CapacityLimit Object
2.2.1.1.	CapacityLimit Types
2.2.1.2.	CapacityLimitTelemetrySource Object
2.2.2.	CapacityLimit Object Serialization
3.	IANA Considerations
3.1.	CDNI Payload Types
3.1.1.	CDNI FCI.Telemetry Payload Type
3.1.2.	CDNI FCI.CapacityLimits Payload Type
3.2.	CDNI Telemetry Source Types Registry
3.2.1.	CDNI Generic Telemetry Source Type
3.3.	CDNI Capacity Limit Types Registry
4.	Security Considerations
5.	References
5.1.	Normative References
5.2.	Informative References
	Acknowledgements
	Authors' Addresses

1. Introduction

While delegating traffic from an upstream CDN (uCDN) to a downstream CDN (dCDN), it is important to ensure that an appropriate amount of traffic is delegated. To achieve that, this specification defines a feedback mechanism to inform the delegator how much traffic may be delegated. The traffic level information provided by that interface will be consumed by services, such as a request router, to inform that service's traffic delegation decisions. The provided information is advisory and does not represent a guarantee, commitment, or reservation of capacity.

This document defines and registers CDNI Payload Types (as defined in Section 7.1 of [RFC8006]). These Payload Types are used for Capability Objects, which are added to those defined in Section 4 of [RFC8008].

1.1. Terminology

The following term is used throughout this document:

CDN: Content Delivery Network

Additionally, this document reuses the terminology defined in [RFC6707]. Specifically, the following CDNI acronyms are used:

uCDN: upstream CDN

dCDN: downstream CDN

1.2. 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.

1.3. Objectives

To enable information exchange between a uCDN and a dCDN regarding acceptable levels of traffic delegation, the following process has been defined:

In normal operation, a uCDN will communicate with a dCDN, via an interface, to collect and understand any limits that a dCDN has set forth for traffic delegation from a uCDN. These limits will come in the form of metrics such as bits per second, requests per second, etc. These limits can be thought of as Not to Exceed (NTE) limits.

The dCDN should provide access to a Telemetry Source of near real-time metrics that the uCDN can use to track current usage. The uCDN should compare its current usage to the limits the dCDN has put forth and adjust traffic delegation decisions accordingly to keep current usage under the specified limits.

In summary, the dCDN will inform the uCDN of the amount of traffic that may be delegated. Additionally, it will provide a Telemetry Source aligned with this limit, allowing the uCDN to monitor its current usage against the advertised value. Having a limit and a corresponding Telemetry Source creates an unambiguous definition understood by both parties.

Limits that are communicated from the dCDN to the uCDN should be considered valid based on the Time to Live (TTL) provided by a mechanism of the underlying transport, e.g., an HTTP Cache-Control header. The intention is that the limits would have a long-lived TTL and would represent a reasonable peak utilization limit that the uCDN should target. If the underlying transport does not provide a mechanism for the dCDN to communicate the TTL of the limits, the TTL should be communicated through an out-of-band mechanism agreed upon between the dCDN and uCDN.

2. CDNI Additional Capability Objects

Section 5 of [RFC8008] describes the FCI Capability Advertisement Object, which contains a CDNI Capability Object as well as the capability-type (a CDNI Payload Type). The section also defines the Capability Objects per such type. Below, we define two additional Capability Objects.

Note: In the following sections, the term "mandatory-to-specify" is used to convey which properties MUST be included when serializing a given capability object. When mandatory-to-specify is defined as a "Yes" for an individual property, it means that if the object containing that property is included in an FCI message, then the mandatory-to-specify property MUST be included.

2.1. Telemetry Capability Object

The Telemetry Capability Object advertises a list of Telemetry Sources made available to the uCDN by the dCDN. In this document, telemetry data is being defined as near real-time aggregated metrics of dCDN utilization, such as bits per second egress, and is specific to the uCDN and dCDN traffic delegation relationship.

Telemetry data is uniquely defined by a source ID, a metric name, and the footprints that are associated with an FCI Capabilities advertisement. When defining a CapacityLimit, the meaning of a limit might be ambiguous if the uCDN and dCDN are observing telemetry via different data sources. A dCDN-provided Telemetry Source that both parties reference serves as a non-ambiguous metric for use when comparing current usage to a limit.

Telemetry data is important for making informed traffic delegation decisions. Additionally, it is essential in providing visibility of traffic that has been delegated. In situations where there are multiple CDN delegations, a uCDN will need to aggregate the usage

information from any dCDNs to which it delegated when asked to provide usage information, otherwise the traffic may seem unaccounted for.

Example: A Content Provider delegates traffic directly to a uCDN, and that uCDN delegates that traffic to a dCDN. When the Content Provider polls the uCDN telemetry interface, any of the traffic the uCDN delegated to the dCDN would become invisible to the Content Provider, unless the uCDN aggregates the dCDN telemetry with its own metrics.

Property: sources

Description: Telemetry Sources made available to the uCDN.

Type: A JSON array of Telemetry Source objects (see Section 2.1.1).

Mandatory-to-Specify: Yes

2.1.1.1. Telemetry Source Object

The Telemetry Source Object is made of an associated type, a list of exposed metrics, and type-specific configuration data.

Property: id

Description: An identifier of a telemetry source. The ID string assigned to this Telemetry Source MUST be unique across all Telemetry Source objects in the advertisement containing this Telemetry Source Object. The ID string MUST remain consistent for the same source reference across advertisements.

Type: String

Mandatory-to-Specify: Yes

Property: type

Description: A valid Telemetry Source Type (see Section 2.1.1.1).

Type: String

Mandatory-to-Specify: Yes

Property: metrics

Description: The metrics exposed by this source.

Type: A JSON array of Telemetry Source Metric Objects (see Section 2.1.1.2).

Mandatory-to-Specify: Yes

Property: configuration

Description: A source-specific representation of the Telemetry Source configuration. For the generic source type, this configuration format is defined as out-of-band. For other types, the configuration format will be specified in a yet-to-be-defined telemetry interface specification. The goal of this element is to allow for forward compatibility with a formal telemetry interface.

Type: A JSON object, the structure of which is specific to the Telemetry Source and outside the scope of this document.

Mandatory-to-Specify: No

2.1.1.1. Telemetry Source Types

At the time of this writing, the "CDNI Telemetry Source Types" registry is limited to a single type: generic (see Table 3 in Section 3.2.1).

2.1.1.2. Telemetry Source Metric Object

The Telemetry Source Metric Object describes the metric to be exposed.

Property: name

Description: An identifier for this metric. This name MUST be unique among metric objects within the containing Telemetry Source. The name MUST remain consistent for the same source reference across advertisements.

Type: String

Mandatory-to-Specify: Yes

Property: time-granularity

Description: The time, in seconds, representing the metric data. For example, a value representing the last 5 minutes would have a time-granularity of 300.

Type: Unsigned Integer

Mandatory-to-Specify: No

Property: data-percentile

Description: The percentile calculation the data represents, i.e., 50 percentile would equate to the median over the time-granularity. Lack of a data-percentile indicates that the data MUST be the mean over the time representation.

Type: Unsigned Integer

Mandatory-to-Specify: No

Property: latency

Description: Time in seconds that the data is behind real-time. This is important to specify to help the uCDN understand how long it might take to reflect traffic adjustments in the metrics.

Type: Unsigned Integer

Mandatory-to-Specify: No

2.1.2. Telemetry Capability Object Serialization

The following shows an example of a Telemetry Capability Object, including two metrics for a source, that is scoped to a footprint.

```
{
  "capabilities": [
    {
      "capability-type": "FCI.Telemetry",
```

```

"capability-value": {
  "sources": [
    {
      "id": "capacity_metrics_region1",
      "type": "generic",
      "metrics": [
        {
          "name": "egress_5m",
          "time-granularity": 300,
          "data-percentile": 50,
          "latency": 1500
        },
        {
          "name": "requests_5m",
          ...
        }
      ]
    }
  ]
},
"footprints": [
  <footprint objects>
]
}
]
}

```

2.2. CapacityLimits Capability Object

The CapacityLimits Capability Object enables the dCDN to specify traffic delegation limits to a uCDN within an FCI Capabilities advertisement. The limits specified by the dCDN will inform the uCDN on how much traffic may be delegated to the dCDN. The limits specified by the dCDN should be considered NTE limits. The limits should be based on near real-time telemetry data that the dCDN provides to the uCDN. In other words, for each limit that is advertised, there should also exist a Telemetry Source that provides current utilization data against the particular advertised limit.

Property: limits

Description: A collection of CapacityLimit Objects.

Type: A JSON array of CapacityLimit Objects (see Section 2.2.1).

Mandatory-to-Specify: Yes

2.2.1. CapacityLimit Object

A CapacityLimit Object is used to represent traffic limits for delegation from the uCDN towards the dCDN. The limit object is scoped to the footprint associated with the FCI Capabilities advertisement encompassing this object. Limits MUST be considered using a logical "AND": A uCDN will need to ensure that all limits are considered rather than choosing only the most specific.

Property: limit-type

Description: The units of maximum-hard and maximum-soft.

Type: String. One of the values listed in Section 2.2.1.1.

Mandatory-to-Specify: Yes

Property: id

Description: Specifies an identifier associated with a limit.
This MAY be used as a relational identifier to a specific CapacityLimit Object. If specified, this identifier MUST be unique among specified identifiers associated with any other CapacityLimit Objects in the advertisement containing this CapacityLimit Object.

Type: String

Mandatory-to-Specify: No

Property: maximum-hard

Description: The maximum unit of capacity that is available for use.

Type: Unsigned Integer

Mandatory-to-Specify: Yes

Property: maximum-soft

Description: A soft limit at which a uCDN SHOULD reduce traffic before hitting the hard limit. This value MUST be less than the value of maximum-hard. If this value is not specified, it is equal to the value of maximum-hard.

Type: Unsigned Integer

Mandatory-to-Specify: No

Property: current

Description: Specifies the current usage value of the limit. It is NOT RECOMMENDED to specify the current usage value inline with the FCI.CapacityLimits advertisements as it will reduce the ability to cache the response, but this mechanism exists for simple use cases where an external Telemetry Source cannot be feasibly implemented. The intended method for providing telemetry data is to reference a Telemetry Source Object (see Section 2.1.1) to poll for the current usage.

Type: Unsigned Integer

Mandatory-to-Specify: No

Property: telemetry-source

Description: The mapping of each particular limit to a specific metric with relevant real-time data provided by a Telemetry Source.

Type: CapacityLimitTelemetrySource object (see Section 2.2.1.2).

Mandatory-to-Specify: No

2.2.1.1. CapacityLimit Types

Below are listed the valid limit-type entries registered in the "CDNI Capacity Limit Types" registry. The values specified here represent the types that were identified as being the most relevant metrics for the purposes of traffic delegation between CDNs.

Capacity Limit Type	Units
---------------------	-------

egress	Bits per second	
+-----+	+-----+	+-----+
requests	Requests per second	
+-----+	+-----+	+-----+
storage-size	Total bytes	
+-----+	+-----+	+-----+
storage-objects	Count	
+-----+	+-----+	+-----+
sessions	Count	
+-----+	+-----+	+-----+
cache-size	Total bytes	
+-----+	+-----+	+-----+

Table 1

2.2.1.2. CapacityLimitTelemetrySource Object

The CapacityLimitTelemetrySource Object refers to a specific metric within a Telemetry Source.

Property: id

Description: Reference to the "id" of a Telemetry Source defined by a Telemetry Capability Object as defined in Section 2.1.

Type: String

Mandatory-to-Specify: Yes

Property: metric

Description: Reference to the "name" property of a metric defined within a Telemetry Source of a Telemetry Capability object.

Type: String

Mandatory-to-Specify: Yes

2.2.2. CapacityLimit Object Serialization

The following shows an example of an FCI.CapacityLimits object.

```
{
  "capabilities": [
    {
      "capability-type": "FCI.CapacityLimits",
      "capability-value": {
        "limits": [
          {
            "id": "capacity_limit_region1",
            "limit-type": "egress",
            "maximum-hard": 50000000000,
            "maximum-soft": 25000000000,
            "telemetry-source": {
              "id": "capacity_metrics_region1",
              "metric": "egress_5m"
            }
          }
        ]
      }
    },
    {
      "footprints": [
        "<footprint objects>"
      ]
    }
  ]
}
```


3. IANA Considerations

3.1. CDNI Payload Types

Per this document, IANA has registered two additional payload types in the "CDNI Payload Types" registry within the "Content Delivery Network Interconnection (CDNI) Parameters" registry group:

Payload Type	Reference
FCI.Telemetry	RFC 9808
FCI.CapacityLimits	RFC 9808

Table 2

3.1.1. CDNI FCI.Telemetry Payload Type

Purpose: The purpose of this Payload Type is to list the supported Telemetry Sources and the metrics made available by each source.

Interface: FCI

Encoding: See Section 2.1.

3.1.2. CDNI FCI.CapacityLimits Payload Type

Purpose: The purpose of this Payload Type is to define Capacity Limits based on utilization metrics corresponding to Telemetry Sources provided by the dCDN.

Interface: FCI

Encoding: See Section 2.2.

3.2. CDNI Telemetry Source Types Registry

IANA has added the following new registry within the "Content Delivery Network Interconnection (CDNI) Parameters" registry group at <https://www.iana.org/assignments/cdni-parameters>:

Registry Name: CDNI Telemetry Source Types

Registry Description: The "CDNI Telemetry Source Types" registry defines the valid values for the "type" property of the Telemetry Source object defined in Section 2.1.1.

Registration Procedure: The registry follows the Specification Required policy as defined in [RFC8126]. The designated expert should consider the following guidelines when evaluating registration requests:

- * The new type definition does not duplicate existing types.
- * The review should verify that the Telemetry Source is applicable to the CDNI use cases and that the description is clear and unambiguous.
- * The registration is applicable for general use and is not proprietary.
- * The "configuration" property has a fully specified object definition with a description of each defined property.

The following value has been registered:

Source Type	Description	Reference
generic	An object that allows for advertisement of generic data sources	RFC 9808

Table 3

3.2.1. CDNI Generic Telemetry Source Type

Purpose: The purpose of this Telemetry Source Type is to provide a source-agnostic telemetry type that may be used for generic Telemetry Source advertisement.

Usage: See Section 2.1.1.

3.3. CDNI Capacity Limit Types Registry

IANA has added the following new registry within the "Content Delivery Network Interconnection (CDNI) Parameters" registry group at <https://www.iana.org/assignments/cdni-parameters>:

Registry Name: CDNI Capacity Limit Types

Registry Description: The "CDNI Capacity Limit Types" registry defines the valid values of the "limit-type" property of a CapacityLimit Object defined in Section 2.2.1.

Registration Procedure: The registry follows the Specification Required policy as defined in [RFC8126]. The designated expert should consider the following guidelines when evaluating registration requests:

- * The new capacity limit-type does not duplicate existing entries.
- * The submission has a defined purpose. The newly defined capacity limit-type should be clearly justified in the context of one or more CDNI use cases.
- * The description of the capacity limit-type is well-documented and unambiguous.

The following values have been registered:

Capacity Limit Type	Units	Reference
egress	Bits per second	RFC 9808
requests	Requests per second	RFC 9808
storage-size	Total bytes	RFC 9808
storage-objects	Count	RFC 9808
sessions	Count	RFC 9808
cache-size	Total bytes	RFC 9808

Table 4

Usage: See Section 2.2.1.1.

4. Security Considerations

This specification is in accordance with the CDNI Request Routing: Footprint and Capabilities Semantics. As such, it is subject to the security and privacy considerations as defined in Section 7 of [RFC8008].

5. References

5.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<https://www.rfc-editor.org/info/rfc2119>>.
- [RFC8008] Seedorf, J., Peterson, J., Previdi, S., van Brandenburg, R., and K. Ma, "Content Delivery Network Interconnection (CDNI) Request Routing: Footprint and Capabilities Semantics", RFC 8008, DOI 10.17487/RFC8008, December 2016, <<https://www.rfc-editor.org/info/rfc8008>>.
- [RFC8126] Cotton, M., Leiba, B., and T. Narten, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 8126, DOI 10.17487/RFC8126, June 2017, <<https://www.rfc-editor.org/info/rfc8126>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

5.2. Informative References

- [RFC6707] Niven-Jenkins, B., Le Faucheur, F., and N. Bitar, "Content Distribution Network Interconnection (CDNI) Problem Statement", RFC 6707, DOI 10.17487/RFC6707, September 2012, <<https://www.rfc-editor.org/info/rfc6707>>.
- [RFC8006] Niven-Jenkins, B., Murray, R., Caulfield, M., and K. Ma, "Content Delivery Network Interconnection (CDNI) Metadata", RFC 8006, DOI 10.17487/RFC8006, December 2016, <<https://www.rfc-editor.org/info/rfc8006>>.
- [SVTA] "Streaming Video Technology Alliance Home Page", <<https://www.svta.org>>.

Acknowledgements

The authors would like to express their gratitude to the members of the Streaming Video Technology Alliance [SVTA] Open Caching Working Group for their guidance, contribution, and review.

Authors' Addresses

Andrew Ryan
Disney Streaming
1211 Avenue of the Americas
New York, NY 10036
United States of America
Email: andrew@andrewnryan.com

Ben Rosenblum

Vecima
4375 River Green Pkwy #100
Duluth, GA 30096
United States of America
Email: ben@rosenblum.dev

Nir B. Sopher
Qwilt
6, Ha'harash
Hod HaSharon 4524079
Israel
Email: nir@apache.org