

# IETF 113 CDNI Working Group Capacity Insights Capability Advertisement Extensions

Mar 2022

# Capacity Insights

What draft are we discussing?

- CDNI Capacity Insights Capability Advertisement Extensions
- draft-ryan-cdni-capacity-insights-extensions-02

# Capacity Insights

What are the goals of the Capacity Extensions being proposed?

- Provide framework for information exchange to facilitate traffic delegation decisions
- Establish limits that are specific to delegation relationship
- Define limits in unambiguous, mutually understood units
- ~~Bidirectional communication between delegation participants~~

# Capacity Insights

How does the communication work?

- FCI.CapacityLimits**
  - Specify the limit of traffic that should be delegated in units, such as Bits per Second, called limit-types
  - References a specific Telemetry source which outlines current usage of a particular limit-type
- FCI.Telemetry**
  - Allows the advertisement of what types of Telemetry sources are supported
  - Initially scoped to generic Telemetry sources, but paves the way for a formal Telemetry interface integration
- MI.RequestedCapacityLimits**
  - Provides a vehicle for the uCDN to ask the dCDN for an update to established Limits

# Capacity Insights – FCI.CapacityLimits

```
{  
  "capabilities": [  
    ...  
    {  
      "capability-type": "FCI.CapacityLimits"  
      "capability-value": {  
        "limits": [  
          {  
            "id" : "region1_total_egress_limit",  
            "limit-type": "egress",  
            "maximum-hard": 5000000000,  
            "maximum-soft": 2500000000,  
            "telemetry-source": {  
              "id": "capacity_metrics_region1",  
              "metric": "egress_5m"  
            },  
            {  
              "id" :  
                "region1_serviceA_egress_limit",  
                "scope": {  
                  "type": "published-host",  
                  "values": [  
                    "serviceA.cdn.example.com"]  
                },  
                "limit-type": "egress",  
                "maximum-hard": 20000000000,  
                "maximum-soft": 10000000000,  
                "current": 15000000000,  
                "telemetry-source": {  
                  "id": "capacity_metrics_region1",  
                  "metric": "egress_service2_5m"  
                }  
              }  
            ]  
          }  
        ]  
      }  
    }  
  ]  
}
```

- limits: collection of limit objects
- limit objects specify the capacity constraints
- default scope of a limit object is tied to the correlating footprint
- limit objects can specify sub scopes within a footprint: defined by scope object
- Scope object defines new type registry of scope type (published host,etc..)
- Each limit references a Telemetry source which provides data with the same scope as the defined limit (i.e Bits per Second for Host X within Footprint Y)
- GOAL: How much can you send, and how can you track how much you are sending

# Capacity Insights – FCI.CapacityLimits – Changes from last draft

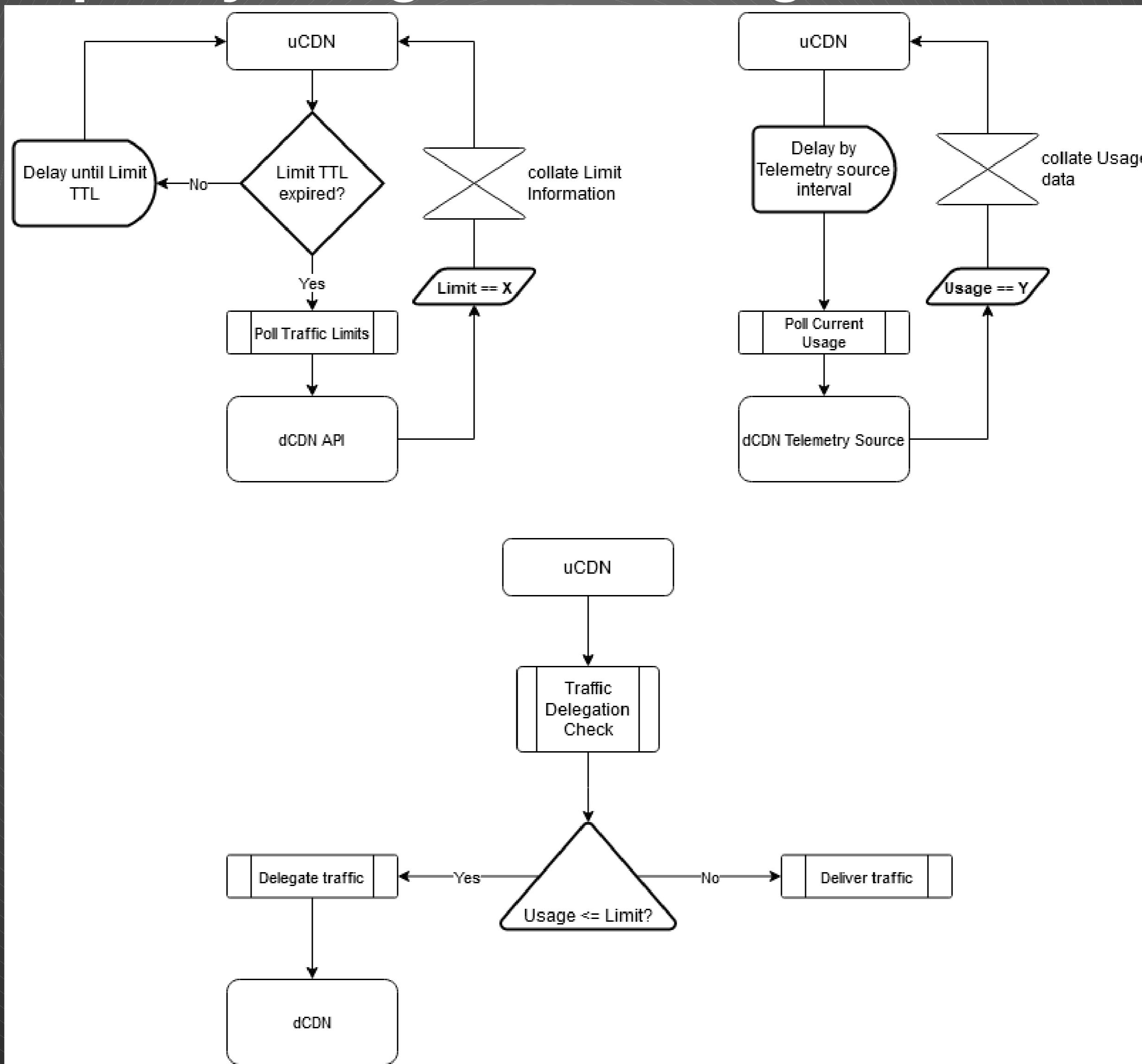
```
{  
  "capabilities": [  
    ...  
    {  
      "capability-type": "FCI.CapacityLimits"  
      "capability-value": {  
        "total-limits": [  
          {  
            "limit-type": "egress",  
            "maximum-hard": 50000000000,  
            "maximum-soft": 25000000000,  
            "telemetry-source": {  
              "id": "capacity_metrics_region1",  
              "metric": "egress_5m"  
            }  
          },  
          ...  
        ],  
        "host-limits": [  
          {  
            "host": "serviceA.cdn.example.com",  
            "limits": [  
              {  
                "limit-type": "egress",  
                "maximum-hard": 20000000000,  
                "maximum-soft": 10000000000,  
                "telemetry-source": {  
                  "id": "capacity_metrics_region1",  
                  "metric": "egress_service2_5m"  
                }  
              ]  
            }  
          ]  
        ]  
      },  
      ...  
    }  
  ]  
},  
...  
{  
  "capabilities": [  
    ...  
    {  
      "capability-type": "FCI.CapacityLimits"  
      "capability-value": {  
        "limits": [  
          {  
            "id": "region1_total_egress_limit",  
            "limit-type": "egress",  
            "maximum-hard": 50000000000,  
            "maximum-soft": 25000000000,  
            "telemetry-source": {  
              "id": "capacity_metrics_region1",  
              "metric": "egress_5m"  
            }  
          },  
          {  
            "id": "region1_serviceA_egress_limit",  
            "scope": {  
              "type": "published-host",  
              "values": [  
                "serviceA.cdn.example.com"  
              ]  
            },  
            "limit-type": "egress",  
            "maximum-hard": 20000000000,  
            "maximum-soft": 10000000000,  
            "telemetry-source": {  
              "id": "capacity_metrics_region1",  
              "metric": "egress_service2_5m"  
            }  
          }  
        ]  
      },  
      ...  
    }  
  ]  
},  
...  
]
```

# Capacity Insights – FCI.Telemetry

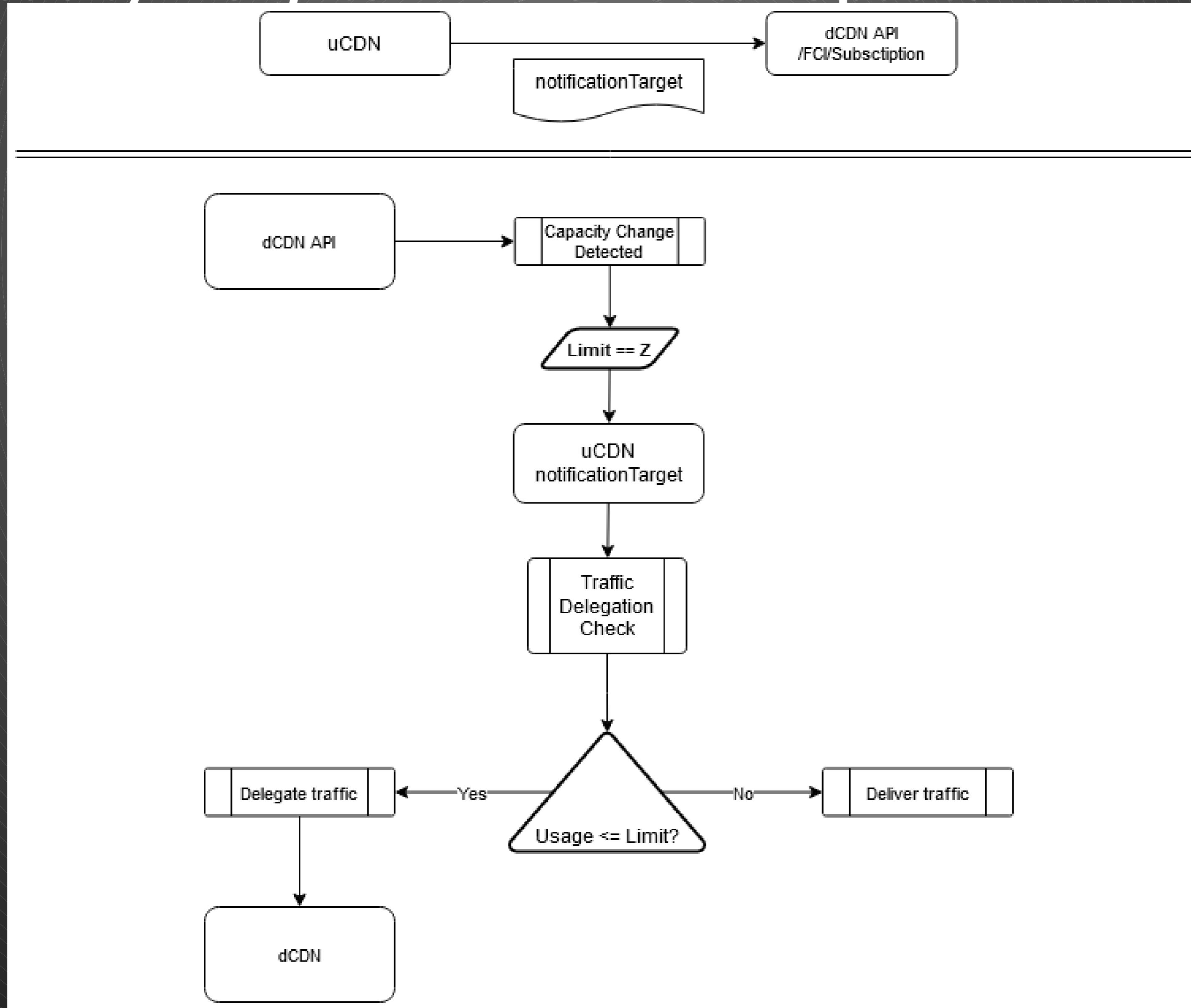
```
"capabilities": [
  {
    "capability-type": "FCI.Telemetry",
    "capability-value": {
      "sources": [
        {
          "id": "capacity_metrics_region1",
          "type": "generic",
          "metrics": [
            {
              "name": "egress_5m",
              "time-granularity": 300,
              "data-percentile": 50,
              "latency": 1500
            },
            {
              "name": "requests_5m",
            },
            ...
          ],
          "configuration": { ... optional ... }
        }
      ]
    }
  ...
]
```

- Lists the supported Telemetry Sources which include an id, type and metrics
- Telemetry data is uniquely identified within a source by id and metrics.name
- FCI.CapacityLimits reference Telemetry sources by id and metrics.name
- Each metric outlines key characteristics of telemetry data such as data latency, granularity, etc..
- Currently type is limited to 'Generic' and is meant to allow for advertisement of ad-hoc/preexisting data sources
- Types will eventually expand to include reference to a formal Telemetry interface
- While this is being defined within the Capacity specification, this is meant to provide a general way to advertise Telemetry data

# Capacity Insights – Polling Workflow



# Capacity Insights – Incremental Update Workflow



# Capacity Insights – MI.RequestedCapacityLimits - Removed

- This object was intended to be used as a means to allow a uCDN to ask a dCDN to reconsider limits
- The usage of this object and the mechanism to transfer this object were not well defined enough to submit for consideration

# Thank you!

---

