

COINRG
Internet-Draft
Intended status: Standards Track
Expires: August 1, 2020

M. McBride
Futurewei
D. Kutscher
Emden University
E. Schooler
Intel
CJ. Bernardos
UC3M
January 29, 2020

Edge Data Discovery for COIN
draft-mcbride-edge-data-discovery-overview-03

Abstract

This document describes the problem of distributed data discovery in edge computing, and in particular for computing-in-the-network (COIN), both the marshalling of data at the outset of a computation and the persistence of the resultant data after the computation. Although the data might originate at the network edge, as more and more distributed data is created, processed, and stored, it becomes increasingly dispersed throughout the network. There needs to be a standard way to find it. New and existing protocols will need to be developed to support distributed data discovery at the network edge and beyond.

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 August 1, 2020.

Copyright Notice

Copyright (c) 2020 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 Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

- [1.](#) Introduction [2](#)
- [1.1.](#) Edge Data [3](#)
- [1.2.](#) Background [3](#)
- [1.3.](#) Requirements Language [4](#)
- [1.4.](#) Terminology [4](#)
- [2.](#) Edge Data Discovery Problem Scope [4](#)
- [2.1.](#) A Cloud-Edge Continuum [5](#)
- [2.2.](#) Types of Edge Data [6](#)
- [2.2.1.](#) Example Meta Data [7](#)
- [3.](#) Scenarios for Discovering Edge Data Resources [8](#)
- [4.](#) Edge Data Discovery [9](#)
- [4.1.](#) Types of Discovery [9](#)
- [4.2.](#) Naming the Data [9](#)
- [5.](#) Use Cases of edge data discovery [10](#)
- [6.](#) IANA Considerations [10](#)
- [7.](#) Security Considerations [11](#)
- [8.](#) Acknowledgement [11](#)
- [9.](#) Normative References [11](#)
- Authors' Addresses [11](#)

[1.](#) Introduction

Edge computing is an architectural shift that migrates Cloud functionality (compute, storage, networking, control, data management, etc.) out of the back-end data center to be more proximate to the IoT data being generated and analyzed at the edges of the network. Edge computing provides local compute, storage and connectivity services, often required for latency- and bandwidth-sensitive applications. Thus, Edge Computing plays a key role in verticals such as Energy, Manufacturing, Automotive, Video

Surveillance, Retail, Gaming, Healthcare, Mining, Buildings and Smart Cities.

[1.1.](#) Edge Data

Edge computing is motivated at least in part by the sheer volume of data that is being created by endpoint devices (sensors, cameras, lights, vehicles, drones, wearables, etc.) at the very network edge and that flows upstream, in a direction for which the network was not originally designed. In fact, in dense IoT deployments (e.g., many video cameras are streaming high definition video), where multiple data flows collect or converge at edge nodes, data is likely to need transformation (transcoded, subsampled, compressed, analyzed, annotated, combined, aggregated, etc.) to fit over the next hop link, or even to fit in memory or storage. Note also that the act of performing compute on the data creates yet another new data stream! Preservation of the original data streams are needed sometimes but not always.

In addition, data may be cached, copied and/or stored at multiple locations in the network on route to its final destination. With an increasing percentage of devices connecting to the Internet being mobile, support for in-the-network caching and replication is critical for continuous data availability, not to mention efficient network and battery usage for endpoint devices.

Additionally, as mobile devices' memory/storage fill up, in an edge context they may have the ability to offload their data to other proximate devices or resources, leaving a bread crumb trail of data in their wakes. Therefore, although data might originate at edge devices, as more and more data is continuously created, processed and stored, it becomes increasingly dispersed throughout the physical world (outside of or scattered across managed local data centers), increasingly isolated in separate local edge clouds or data silos. Thus there needs to be a standard way to find it. New and existing protocols will need to be identified/developed/enhanced for these purposes. Being able to discover distributed data at the edge or in the middle of the network - will be an important component of Edge computing.

[1.2.](#) Background

Several IETF T2T RG Edge Computing discussions have been held over the last couple years, a comparative study on the definition of Edge computing was presented in multiple sessions in T2T RG in 2018 and an Edge Computing I-D was submitted early 2019. An IETF BEC (beyond edge computing) effort has been evaluating potential gaps in existing edge computing architectures. Edge Data Discovery is one potential

gap that was identified and that needs evaluation and a solution. The newly proposed COIN RG highlights the need for computations in the network to be able to marshal potentially distributed input data and to handle resultant output data, i.e., its placement, storage and/or possible migration strategy.

1.3. 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](#)].

1.4. Terminology

- o Edge: The edge encompasses all entities not in the back-end cloud. The device edge is the boundary between digital and physical entities in the last mile network. Sensors, gateways, compute nodes are included. The infrastructure edge includes equipment on the network operator side of the last mile network including cell towers, edge data centers, cable headends, POPs, etc. See Figure 1 for other possible tiers of edge clouds between the device edge and the back-end cloud data center.
- o Edge Computing: Distributed computation that is performed near the network edge, where nearness is determined by the system requirements. This includes high performance compute, storage and network equipment on either the device or infrastructure edge.
- o Edge Data Discovery: The process of finding required data from edge entities, i.e., from databases, files systems, device memory that might be physically distributed in the network, and providing access to it logically as if it were a single unified source, perhaps through its namespace, that can be evaluated or searched.
- o ICN: Information Centric Networking. An ICN-enabled network routes data by name (vs address), caches content natively in the network, and employs data-centric security. Data discovery may require that data be associated with a name or names, a series of descriptive attributes, and/or a unique identifier.

2. Edge Data Discovery Problem Scope

Our focus is on how to define and scope the edge data discovery problem. This requires some discussion of the evolving definition of the edge as part of a cloud-to-edge continuum and in turn what is meant by edge data as well as the meta-data about edge data.

[2.1.](#) A Cloud-Edge Continuum

Although Edge Computing data typically originates at edge devices, there is nothing that precludes edge data from being created anywhere in the cloud-to-edge computing continuum (Figure 1). New edge data may result as a byproduct of computation being performed on the data stream anywhere along its path in the network. For example, infrastructure edges may create new edge data when multiple data streams converge upon this aggregation point and require transformation (e.g., to fit within the available resources, to smooth raw measurements to eliminate high-frequency noise, to obfuscate data for privacy).

An assumption is that all data will have associated policies (default, inherited or configured) that describe access control permissions. Consequently, the discoverability of data will be a function of who or what has requested access. In other words, the discoverable view into the available data will be limited to those who are authorized. Discovering edge data that is exclusively private is out of scope of this document, the assumption being that there will be some edge clouds that do not expose or publish the availability of their data. Although edge data may be sent to the back-end cloud as needed, there is nothing that precludes it from being discoverable if the cloud offers it as public.

Initially our focus is on discovery of edge data that resides at the Device Edge and the Infrastructure Edge.

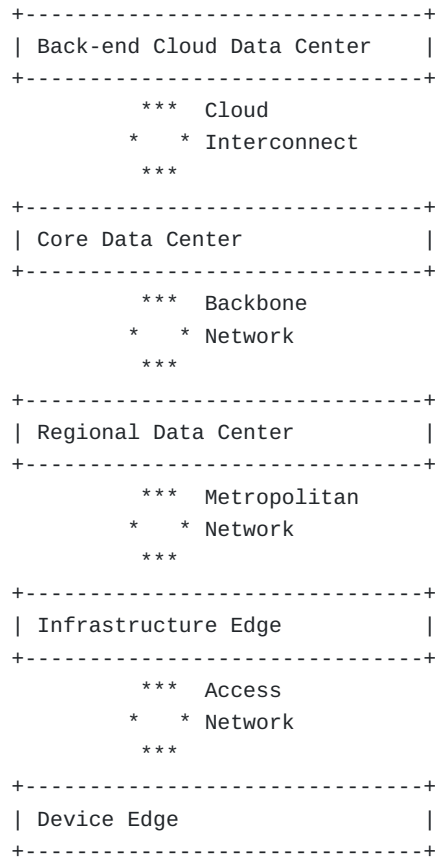


Figure 1: Cloud-to-edge computing continuum

2.2. Types of Edge Data

Besides classically constrained IoT device sensor and measurement data accumulating throughout the edge computing infrastructure, edge data may also take the form of higher frequency and higher volume streaming data (from a continuous sensor or from a camera), meta data (about the data), control data (regarding an event that was triggered), and/or an executable that embodies a function, service, or any other piece of code or algorithm. Edge data also could be created after multiple streams converge at an edge node and are processed, transformed, or aggregated together in some manner.

Regardless of edge data type, a key problem in the Cloud-Edge continuum is that data is often kept in silos. Meaning, data is often sequestered within the Edge where it was created. A goal of this discussion is to consider the prospect that different types of edge data will be made accessible across disparate edges, for example to enable richer multi-modal analytics. But this will happen only if

data can be described, searched and discovered across heterogeneous edges in a standard way. Having a mechanism to enable granular edge data discovery is the problem that needs solving either with existing or new protocols. The mechanisms shouldn't care to which flavor cloud or edge the request for data discovery is made.

[2.2.1.](#) Example Meta Data

SFC Data and meta-data discovery

Service function chaining (SFC) allows the instantiation of an ordered set of service functions and subsequent "steering" of traffic through them. Service functions provide a specific treatment of received packets, therefore they need to be known so they can be used in a given service composition via SFC. So far, how the SFs are discovered and composed has been out of the scope of discussions in IETF. While there are some mechanisms that can be used and/or extended to provide this functionality, work needs to be done. An example of this can be found in [[I-D.bernardos-sfc-discovery](#)].

In an SFC environment deployed at the edge, the discovery protocol may also need to make available the following meta-data information per SF:

- o Service Function Type, identifying the category of SF provided.
- o SFC-aware: Yes/No. Indicates if the SF is SFC-aware.
- o Route Distinguisher (RD): IP address indicating the location of the SF(I).
- o Pricing/costs details.
- o Migration capabilities of the SF: whether a given function can be moved to another provider (potentially including information about compatible providers topologically close).
- o Mobility of the device hosting the SF, with e.g. the following sub-options:
 - Level: no, low, high; or a corresponding scale (e.g., 1 to 10).
 - Current geographical area (e.g., GPS coordinates, post code).
 - Target moving area (e.g., GPS coordinates, post code).
- o Power source of the device hosting the SF, with e.g. the following sub-options:

Battery: Yes/No. If Yes, the following sub-options could be defined:

Capacity of the battery (e.g., mmWh).

Charge status (e.g., %).

Lifetime (e.g., minutes).

Discovery of resources in an NFV environment: virtualized resources do not need to be limited to those available in traditional data centers, where the infrastructure is stable, static, typically homogeneous and managed by a single admin entity. Computational capabilities are becoming more and more ubiquitous, with terminal devices getting extremely powerful, as well as other types of devices that are close to the end users at the edge (e.g., vehicular onboard devices for infotainment, micro data centers deployed at the edge, etc.). It is envisioned that these devices would be able to offer storage, computing and networking resources to nearby network infrastructure, devices and things (the fog paradigm). These resources can be used to host functions, for example to offload/complement other resources available at traditional data centers, but also to reduce the end-to-end latency or to provide access to specialized information (e.g., context available at the edge) or hardware. Similar to the discovery of functions, while there are mechanisms that can be reused/extended, there is no complete solution yet defined. An example of work in this area is [\[I-D.bernardos-intarea-vim-discovery\]](#)."

3. Scenarios for Discovering Edge Data Resources

1. A set of data resources appears (e.g., a mobile node hosting data joins a network) and they want to be discovered by an existing but possibly virtualized and/or ephemeral data directory infrastructure.
2. A device wants to discover data resources available at or near its current location. As some of these resources may be mobile, the available set of edge data may vary over time.
3. A device wants to discover to where best in the edge infrastructure to opportunistically upload its data, for example if a mobile device wants to offload its data to the infrastructure (for greater data availability, battery savings, etc.)

[4. Edge Data Discovery](#)

How can we discover data on the edge and make use of it? There are proprietary implementations that collect data from various databases and consolidate it for evaluation. We need a standard protocol set for doing this data discovery, on the device or infrastructure edge, in order to meet the requirements of many use cases. We will have terabytes of data on the edge and need a way to identify its existence and find the desired data. A user requires the need to search for specific data in a data set and evaluate it using their own tools. The tools are outside the scope of this document, but the discovery of that data is in scope.

[4.1. Types of Discovery](#)

There are many aspects of discovery and many different protocols that address each aspect.

Discovery of new devices added to an environment. Discovery of their capabilities/services in client/server environments. Discovery of these new devices automatically. Discovering a device and then synchronizing the device inventory and configuration for edge services. There are many existing protocols to help in this discovery: UPnP, mDNS, DNS-SD, SSDP, NFC, XMPP, W3C network service discovery, etc.

Edge devices discover each other in a standard way. We can use DHCP, SNMP, SMS, COAP, LLDP, and routing protocols such as OSPF for devices to discovery one another.

Discovery of link state and traffic engineering data/services by external devices. BGP-LS is one solution.

The question is if one or more of these protocols might be a suitable contender to extend to support edge data discovery?

[4.2. Naming the Data](#)

Named Data Networking (NDN) is one of five research projects funded by the U.S. National Science Foundation under its Future Internet Architecture Program. NDN has its roots in an earlier project, Content-Centric Networking (CCN), which Van Jacobson started at Xerox PARC around the time of his Google talk, to turn his architecture vision into a running prototype (see also his CoNEXT 2009 paper and especially Jacobsons ACM Queue interview). The motivation is the mis-match of today's Internet architecture and its usage. Today we build, support, and use Internet applications and services on top of an extremely capable architecture not designed to support them. What

if we had an architecture designed to support them? Specifically, today's IP packets can name only endpoints of conversations (IP addresses) at the network layer. What if we generalize this layer to name any information (or content), not just endpoints? We make it easier to develop, manage, secure, and use our networks. NDN can be applied to edge data discovery to make it much easier to extract data and meta-data by naming it. If data was named we would be able to discover the appropriate data simply by its name.

5. Use Cases of edge data discovery

1. Autonomous Vehicles

Autonomous vehicles rely on the processing of huge amounts of complex data in real-time for fast and accurate decisions. These vehicles will rely on high performance compute, storage and network resources to process the volumes of data they produce in a low latency way. Various systems will need a standard way to discover the pertinent data for decision making

2. Video Surveillance

The majority of the video surveillance footage will remain at the edge infrastructure (not sent to the cloud data center). This footage is coming from vehicles, factories, hotels, universities, farms, etc. Much of the video footage will not be interesting to those evaluating the data. A mechanism, set of protocols perhaps, is needed to identify the interesting data at the edge. What constitutes interesting will be context specific, e.g., video frames with a car in it, a backyard nocturnal creature in it, a person or bicyclist or etc. Interesting video data may be stored longer in storage systems at the very edge of the network or in flight in networking equipment further away from the device edge.

3. Elevator Networks

Elevators are one of many industrial applications of edge computing. Edge equipment receives data from 100's of elevator sensors. The data coming into the edge equipment is vibration, temperature, speed, level, video, etc. We need the ability to identify where the data we need to evaluate is located.

6. IANA Considerations

N/A

7. Security Considerations

Security considerations will be a critical component of edge data discovery particularly as intelligence is moved to the extreme edge where data is to be extracted.

8. Acknowledgement

9. Normative References

[I-D.bernardos-intarea-vim-discovery]

Bernardos, C. and A. Mourad, "IPv6-based discovery and association of Virtualization Infrastructure Manager (VIM) and Network Function Virtualization Orchestrator (NFVO)", [draft-bernardos-intarea-vim-discovery-02](#) (work in progress), August 2019.

[I-D.bernardos-sfc-discovery]

Bernardos, C. and A. Mourad, "Service Function discovery in fog environments", [draft-bernardos-sfc-discovery-03](#) (work in progress), September 2019.

[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

Mike McBride
Futurewei

Email: michael.mcbride@futurewei.com

Dirk Kutscher
Emden University

Email: ietf@dkutscher.net

Eve Schooler
Intel

Email: eve.m.schooler@intel.com

URI: <http://www.eveschooler.com>

Carlos J. Bernardos
Universidad Carlos III de Madrid
Av. Universidad, 30
Leganes, Madrid 28911
Spain

Phone: +34 91624 6236

Email: cjbc@it.uc3m.es

URI: <http://www.it.uc3m.es/cjbc/>