Data Discovery

draft-mcbride-data-discovery-problem-statement
draft-mcbride-edge-data-discovery-overview

Mike McBride - Futurewei
Dirk Kutscher - Emden University
Eve Schooler - Intel
Carlos J Bernardos - UC3M
Diego Lopez - Telefonica
Xavier de Foy - InterDigital

What's the problem?

Data increasingly is being created, cached, copied, transformed at multiple locations in the network on route to its final destination

Would like to discover
(1) where pockets of data exist and
(2) where specific data objects are located

Would like to be able to locate distributed data in an open standardized way

What is Data?

- **Data:** statistics, measurements, temperature, location, metadata, health, etc

- **Bag of bits:**
  - **Program**: applets, graphics, games, spreadsheets, database systems, browsers, etc
  - **Service**: firewalls, load-balancers, spam filters, header manipulators, etc

- **Meta-data:**
  - **Resource**: CPU, memory, etc.

Data Discovery is fundamental to COIN

Computation = a function performed on Data that generates Data

Input → Function → Output

- **Input data** – marshall from somewhere
- **Function** – the method, algorithm, executable – sometimes determined on the fly, based on context
- **Output data** - may need to be placed somewhere after the computation, if persistent (for re-use)

Why COIN RG?

Pros

- RG is a good staging ground to mature ideas
- Data Discovery is an early stage idea - no specific protocol or architectural recommendations as yet
- Data Discovery is motivated by COIN (& mobility, net dynamics)

Cons

- Seems off topic - based on group charter
- Seems less off topic – if following progress of discussion & drafts
- Scope of docs have grown vs narrowed – relevance blurred some
  - One outcome: better characterize what needs discovery in COIN

What's next?

- Request WG adoption

BACKUP

Table of Contents - 05

Table of Contents

1. Introduction 2
   1.1. Edge Data 3
   1.2. Background 4
   1.3. Requirements Language 4
   1.4. Terminology 4
2. Edge Data Discovery Problem Scope 5
   2.1. A Cloud-Edge Continuum 5
   2.2. Types of Edge Data 6
3. Edge Scenarios Requiring Data Discovery 7
4. Edge Data Discovery 7
   4.1. Types of Discovery 7
   4.2. Early Stage of Discovery 8
   4.3. Naming the Data 8
5. Use Cases of Edge Data Discovery 10
   5.1. Autonomous Vehicles 10
   5.2. Video Surveillance 10
   5.3. Elevator Networks 10
   5.4. Service Function Chaining 11
   5.5. Ubiquitous Witness 12
6. IANA Considerations 13
7. Security Considerations 13
8. Acknowledgement 14
9. Normative References 14
Authors' Addresses 15

Slides from IETF 108

What's the Problem?

- Increasing #s of devices and sensors generate a torrent of data
  - at the network Edge that flows upstream
- Sometimes that data must be processed/transformed → new data!
  - e.g., transcoded, subsampled, compressed, analyzed, annotated, combined, aggregated, et cetera
- In addition, (transformed) data may be cached/stored at multiple locations in the network on route to its final destination
- **As more distributed data is created, processed and stored, it becomes increasingly dispersed**
  - Throughout the network
- There needs to be a standard way – or ways - to find it!
  - New and existing protocols may need to be identified/developed/enhanced for distributed data discovery at the network edge …and beyond

How does this relate to COIN?

COIN requires data input and often results in data output:

- From where does COIN expect the data to come? To where does it expect it to be cached or to flow afterwards?
- How should the availability of data be exposed, where appropriate, while at the same time its privacy preserved?
- How to ensure COIN protocols comprehend the Edge context where data may not be movable (because of its abundance)?

Scenarios Requiring Discovery of Edge Data

1. A set of data resources appears (e.g., a mobile node joins the net)
   - want to be discoverable by an existing but possibly virtualized and/or ephemeral data directory infrastructure.

2. A device wants to discover data resources near its current location
   - because some resources may be mobile, asleep, or only intermittently connected, the available set of edge data may vary over time.

3. A device wants to discover where best in the edge infrastructure to opportunistically upload/migrate its data
   - if a mobile device wants to offload its data to the infrastructure (for greater data availability, battery savings, safe keeping, etc.).
   - a network element is running out of space

Table of Contents - 04

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 5
  2.1. A Cloud-Edge Continuum 5
  2.2. Types of Edge Data 6
3. Edge Scenarios Requiring Data Discovery 7
4. Edge Data Discovery 7
  4.1. Types of Discovery 7
  4.2. Naming the Data 8
5. Use Cases of Edge Data Discovery 9
  5.1. Autonomous Vehicles 9
  5.2. Video Surveillance 10
  5.3. Elevator Networks 10
  5.4. Service Function Chaining 10
  5.5. Ubiquitous Witness 12
6. IANA Considerations 13
7. Security Considerations 13
8. Acknowledgement 13
9. Normative References 13
Authors' Addresses 14