Edge Data Discovery for COIN

draft-mcbride-edge-data-discovery-overview-04

Mike McBride (Futurewei)
Dirk Kutscher (Emden University)
Eve M. Schooler (Intel)
Carlos J Bernardos (UC3M)
Diego Lopez (Telefonica)

July 30, 2020

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)?

Table of Contents - NEW

Modifications

- Added new Author
- Addressed detailed edits from Greg Skinner
- Use case updates
  - Added Ubiquitous Witness
    - Contextually-related data, data search, group processing
  - Clarified Service Function Chaining
- Began a very modest Security section
  - Data policies, discoverability a fn of access control, assumptions
- Received comments from Lixia Zhang
  - Data Provenance
  - Definition of Edge computing (not simply moving data closer)

BACKUP

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