Overview of Edge Data Discovery

draft-mcbride-edge-data-discovery-overview-01

Mike McBride (Huawei)
Dirk Kutscher (Emden University)
Eve M. Schooler (Intel)
Carlos J Bernardos (UC3M)

March 2019

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 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

Clarified definitions:
- Edge computing
- Named Data Networking
- Edge data locations

e.g., streaming sensor or measurement data, streaming media, meta-data, functions/services, bag of bits….

Focused on Data Discovery (vs general discovery problem)

Feedback – main input

- Edge data discovery – crisper definition of problem needed
- Broader Edge data life-cycle management problem
  - Discover, Search, Access, …**Compute/Transform**, …Pin/Place, Migrate, Expire, Secure, Preserve-privacy, Support mobility, etc.
- Include device in the Cloud-to-Edge continuum?
  - Treat device separately, because of security, privacy, mobility?
    - Devices are authoritative re the data they have
  - What about discovery in a P2P manner?

- Convert NDN discussion into an ICN discussion
- Better integrate the section on service function chaining (meta) data
- Security section needs more serious thought

Feedback – smaller issues

- IoT Data vs any Data
- Data is needed by analysis to make local and/or low-latency decisions - for predictive maintenance? emergency services?
- Video analytics as a vertical segment?
- Tease apart caching and replication, as well as caching and storage
- Placement of business case for data economy
- Not all Clouds silo data. Not all Edges expose data.

Next Steps

- Address all feedback in next version -02
- Capture interplay of COIN & the broader Edge data life-cycle
  - Investigate more completely the SoTA, requirements, considerations
  - Begin by drilling into references & relevance of current discovery protocols
- Determine if existing protocols could/should be extended
  - Involve CoAP, DNS SD, W3C Thing Directory?
  - resource directories for named data?
- Determine if a new discovery protocol may be needed