Proposed RG: COIN Computing in the Network

Jeffrey He, Marie-José Montpetit, Eve M. Schooler

IETF 104 - Prague

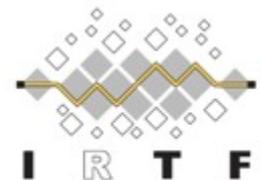*March 28th, 2019*

Note Well (IRTF)

This is a reminder of IETF policies in effect on various topics such as patents or code of conduct. It is only meant to point you in the right direction. Exceptions may apply. The IETF's patent policy and the definition of an IETF "contribution" and "participation" are set forth in BCP 79; please read it carefully. The IRTF follows the IETF's policies.

As a reminder:

- **By participating in the IRTF, you agree to follow IETF processes and policies.**
- **If you are aware that any IETF or IRTF contribution is covered by patents or patent applications that are owned or controlled by you or your sponsor, you must disclose that fact, or not participate in the discussion.**
- **As a participant in or attendee to any IRTF activity you acknowledge that written, audio, video, and photographic records of meetings may be made public.**
- **Personal information that you provide to IRTF will be handled in accordance with the IETF Privacy Statement.**
- **As a participant or attendee, you agree to work respectfully with other participants; please contact the ombudsteam (https://www.ietf.org/contact/ombudsteam/) if you have questions or concerns about this.**

Definitive information is in the documents listed below and other IETF BCPs. For advice, please talk to the RG Chairs or IRTF Chair:

- **BCP 9 (Internet Standards Process)**
- **BCP 25 (Working Group processes)**
- **BCP 25 (Anti-Harassment Procedures)**
- **BCP 54 (Code of Conduct)**
- **BCP 78 (Copyright)**
- **BCP 79 (Patents, Participation)**
- **https://www.ietf.org/privacy-policy/ (Privacy Policy)**

I R T F

Administrivia

- **Wiki** – documents, proposed charter, today's slides
  https://trac.ietf.org/trac/irtf/wiki/coin

- **Mailing list** – coin@irtf.org

- **Remote participation**
  - Meeting link: http://www.meetecho.com/ietf104/coinrg/

Agenda

00   Review of the agenda (Chairs - 5 minutes)

01   Current charter - context, motivation, objectives (Chairs/All- 15 minutes)

Presentations

02   Industrial Networks (Klaus Wehrle - 10 minutes)

03   Edge data discovery (Michael McBride/Eve Schooler- 10 minutes)

04   A keyword-based naming for in-network computing (Borje Ohlman – 10 minutes)

05   IoT Edge Computing/T2T RG Liaison (Matthias Kovatsch – 5 minutes)

06   Computing Service Providers (Noa Zilberman (remote - 15 minutes)

Discussion

07   Next steps/milestones/group directions (Chairs/All 20 minutes)

Existing IDs

- draft-he-managed-networks-00
  - https://datatracker.ietf.org/doc/draft-he-coin-managed-networks
- draft-montpetit-coin-XR-02
  - https://tools.ietf.org/html/draft-montpetit-coin-xr-02
- draft-mcbride-edge-data-discovery-01
  - https://tools.ietf.org/pdf/draft-mcbride-edge-data-discovery-overview-01.pdf

Collaborators welcome and new IDs encouraged!

COIN

- Our goal:

    Foster research in computing in the network to improve performance for networks and applications

- Focus:
    - Architectures
    - Protocols
    - Real-world use cases, applications, work in progress

- Charter:
    - We have modified the charter after Bangkok the meeting, based on the feedback form the mailing list (thanks Dave Oran, Dirk Kutscher, Laurent Ciavaglia and others).
    - See the wiki (https://trac.ietf.org/trac/irtf/wiki/coin)

Objectives

- Systematically look at different instantiations of COIN in the cloud-to-edge-to-device computing continuum:

  - Are there common principles, abstractions, assumptions and mechanisms that can be applied across this range of different types of computing/networking applications/approaches?
  - What are best practices and relevant considerations for computing/networking systems, in particular with respect to previous efforts in active networking?
  - For a network with non-forwarding functions, at what level of abstraction must the programmable data plane live?
  - What is the impact of in-network functions on end-to-end transport protocols and security?
  - What are the incentives for the network to add new computing capabilities and resources in an open eco-system?

Cloud-to-Edge Computing Continuum

Discussion/Agreement since IETF 103:

The need for a cloud-to-edge computing continuum goes hand in hand with the need to support computing-in-the-network (COIN)
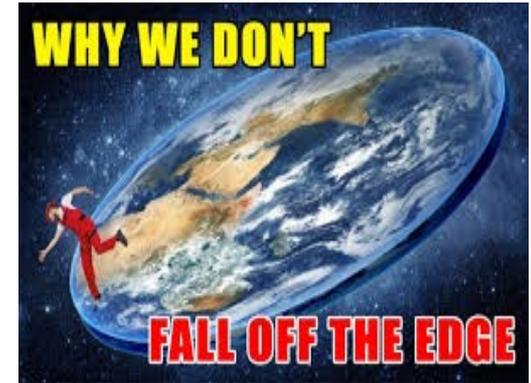
What does this mean?

in the beginning…

- Many (IoT) applications need a more proximate Cloud (aka Edge Cloud)
  - At first, this meant a closer, scaled-down Cloud data center
  - In effect, migrate cloud functionality (compute, storage, networking, control, data management, etc) to be more proximate to where data is being generated, analyzed/computed upon/transformed and acted upon.

- Has led to a proliferation of Edges
  - Some more proximate than others, including all the way to the edge device.
  - Some are in the likeness of cloud data centers, or server racks in Central Offices, but others are living in much less managed spaces and are distinctly more distributed
  - Some working as a (logical) federation of resources (meaning, not physically co-located), some a single resource

- In the extreme, everything is an Edge… or Cloud… or Data center
  - Thus Edge computing is a stepping stone to Beyond Edge Computing or Fog computing or Ambient Computing etc.

Limitations at the Edge?

- Not every Edge is going to support the full suite of functionality of a back-end Cloud data-center

- Therefore the Cloud is becoming not only decentralized and distributed, but also disaggregated
  - Its services and microservices may be scattered across a series of entities in the network
  - They may or may not federate to act collectively as one Edge, but may offer these services independently

Location, location, location!

Depending on where it is placed...

- **Compute looks categorically different**, e.g., when in

    - **Data centers**, compute happens on (racks of) Servers
        - Key-value store, Consensus
    - **Network switches, firewalls and routers**
        - Packet inspection, Acceleration (ASICs, FPGAs, etc), Crypto
    - **RSUs**
        - High Definition Map updates
        - Stitching upstream video into MPEG-I immersive AR/VR
    - **Cameras**
        - AI/Analytics for object recognition
        - Labelling video streams with meta-data re scene understanding
- **Some compute may be more managed than others**
    - By different types of Providers - e.g., cloud, comms, gaming, streaming media....

FIZZ BEER CO.

FIZZ

FOR SALE

Search ID: rhan1032

WE'LL TAKE IT!

Implications of the Continuum?

There are many places to put Compute in the Network - from Cloud to Edge to Device

- In what form factors, "boxes", and/or infrastructure might we encounter the need for Compute, throughout the continuum?

    - **Stationary:** Racks of servers in "central office" data centers at varying distances from the back-end Cloud data-center, Base stations and APs, Road side units (RSUs), Cameras, Gaming consoles, etc.

    - **Mobile (wireless):** Cars, phones, drones, robots, head-mounted displays (HMDs)

More Implications/Questions

- Where in the HW/FW/SW stack will the Compute expect to be performed?
  - Not all "Network boxes" will perform Network-related functions
  - What does it mean to perform packet processing?
    - Compute may need to be optimized to handle packet content (vs just packet headers)?
  - What does programmability of these Network elements look like?

- For truly improved performance, Network functionality needs to be designed to "play well" not only with Compute, but also with Storage!
  - Networks are becoming more than "just" Networks

- Developers/users don't want to have to specify where exactly COIN will run in the Cloud-to-Edge computing continuum
  - They just want the task requirements to be met
  - And something/somewhere to figure out the placement of the task components (for good-enough, improved, or optimal performance)

- If an abundance of Data originates at Edge devices and flows upstream, Edge aggregator nodes may need to perform Compute to reduce resource usage and/or to manage data implosion

Goals of Today's Presentations

- Explore the COIN continuum & the state-of-the-art
- Define potential topics and use cases for the RG to focus on
- Help to identify what else is needed
- Help to form opinions on the PRG scope
- Identify potential drafts and bring contributors together

Presentations

Discussion

- Reminder: goals of today's presentations
  - Explore the COIN continuum & the state-of-the-art
  - Define potential topics and use cases for the RG to focus on
  - Help to identify what else is needed
  - Help to form opinions on the PRG scope
  - Identify potential drafts and bring contributors together

What's the Commonality?

All Computing in the Network needs to…

- Select a programming paradigm to launch/orchestrate the process
- Marshal resources
  - Such as: Compute, Network, Storage, Data, Telemetry, Control, etc.…
  - Each of which must be able to describe and expose their capabilities
- Meet requirements/constraints/policies
- Adapt to changing conditions
- Leverage telemetry
- Establish strong security/privacy/trust guarantees
- ???

Next Steps

- PRG activities:
  - Articulate potential "gains", problems & challenges
    - For IETF concerns?
  - Study Commonality across Cloud-to-Edge computing continuum
  - Identify (more) real use cases and their requirements
  - Comprehend Programmability models
  - Solicit proposals for architecture, protocols, eco-system solutions

- Foster informal dialog with related groups: *e.g.,*
  - IETF WGs/RGs: T2T, DIN, ICN, DetNet, CoRE, DNSSD, TEEP, RATS
  - External consortia: ETSI MEC, LFEdge/EdgeX, ECC, IIC/OpenFog

- Meetings
  - Potential (virtual) interim meeting ahead of Montreal IETF?
  - Meet at IETF 105 in Montreal (July 20-26)