In-Network Computing for App-Centric Micro-Services

draft-sarathchandra-coin-appcentres-02

C. Sarathchandra, D. Trossen, M. Boniface

COIN RG virtual meeting 07.04.2020

Premise of the Draft

*"*The application-centric deployment of 'Internet' services has increased over the past ten years with many million applications providing user-centric services, executed on increasingly more powerful smartphones that are supported by Internet-based cloud services in distributed data centres, the latter mainly provided by large scale players such as Google, Amazon and alike. *This draft outlines a vision of evolving those data centres towards executing app-centric micro-services; we dub this evolved data centre as an **AppCentre**."*

- Draft outlines use cases and research challenges for this vision

General Structure

Added new use cases and pulled requirements into separate sub-section

Minor revision with more planned in next version

Mobile Application Function Offloading

Main Idea
Push towards *ubiquitous computing* through distributing mobile applications via a micro-service based design, linking into COIN RG and similar activities
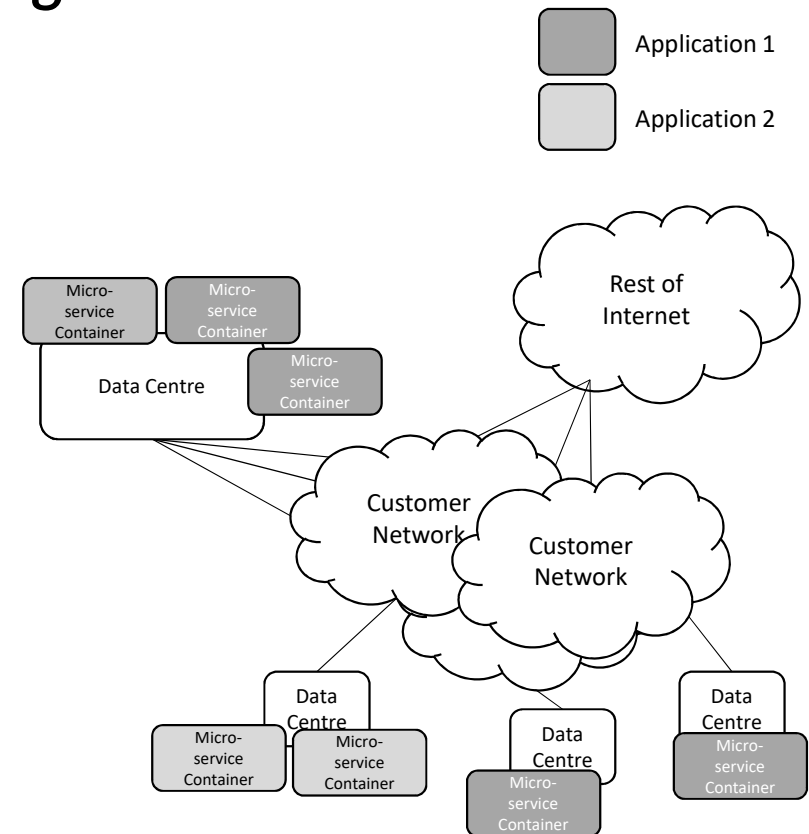
Drivers
Rise of peripherals and the needed collaboration, e.g., multi-surface VR, multi-viewer VR, gaming interaction

Examples
Any mobile app really, e.g., (1) multi-viewing experiences (2) multi-user gaming (3) localized tourist experiences

Current Technology Pain Points
- No native support for HTTP-based micro-services, i.e., *platforms required*
- *Latency*, often caused, e.g., TCP close after every HTTP request and transport handshake
- *Chaining overhead* adds to already existing transport/routing overhead

Application 1

Application 2

Micro-service Container
Micro-service Container
Micro-service Container

Data Centre

Rest of Internet

Customer Network

Customer Network

Data Centre
Micro-service Container
Micro-service Container

Data Centre
Micro-service Container

Data Centre
Micro-service Container

4

Distributed AI

Main Idea
Utilize distribute computing power to increase compute capabilities as well as utilize localized data & reasoning
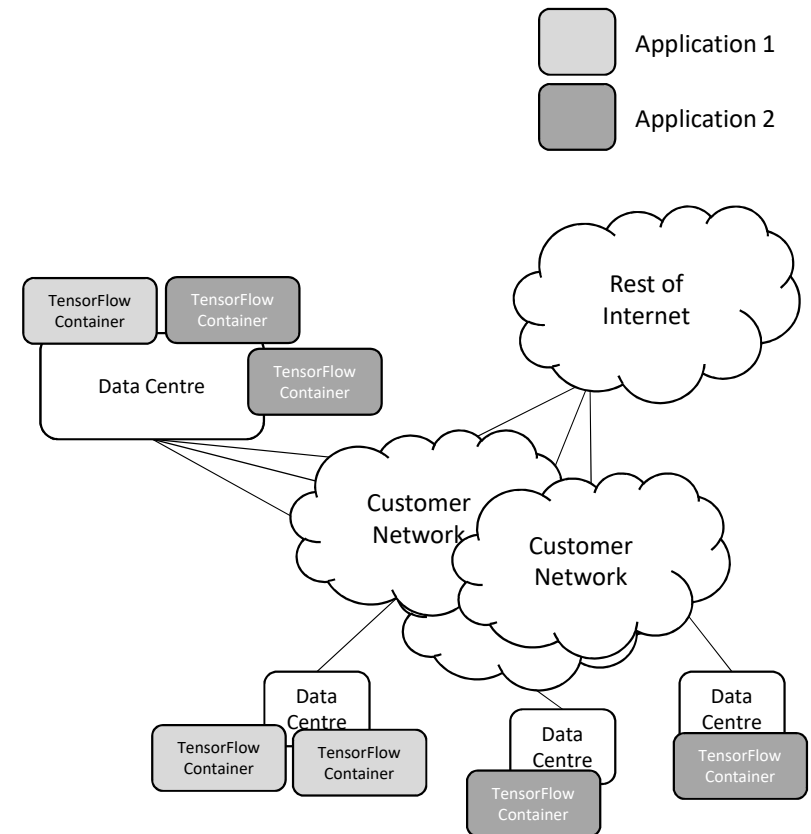
Drivers
(1) compute multiplication (for areas like bio/astro) (2) GDPR driving to localize reasoning over sensitive data

Examples
(1) RAN processing (2) RADAR like applications (3) Large-scale image recognition (4) V2X (5) smart energy

Current Technology Pain Points
- *Latency* since message passing needs to preserve semantics of onboard neural processing communication
- *Chaining* to be done in application with no native platform support
- Integration of *rich processing endpoints*, e.g., base stations

Application 1

Application 2

Rest of Internet

Data Centre

TensorFlow Container

TensorFlow Container

TensorFlow Container

Customer Network

Customer Network

Data Centre

TensorFlow Container

TensorFlow Container

Data Centre

TensorFlow Container

Data Centre

TensorFlow Container

Content Delivery Networks

Main Idea
Utilize multicast opportunity in path-based forwarding to improve on distribution within CDN and towards customers

Drivers
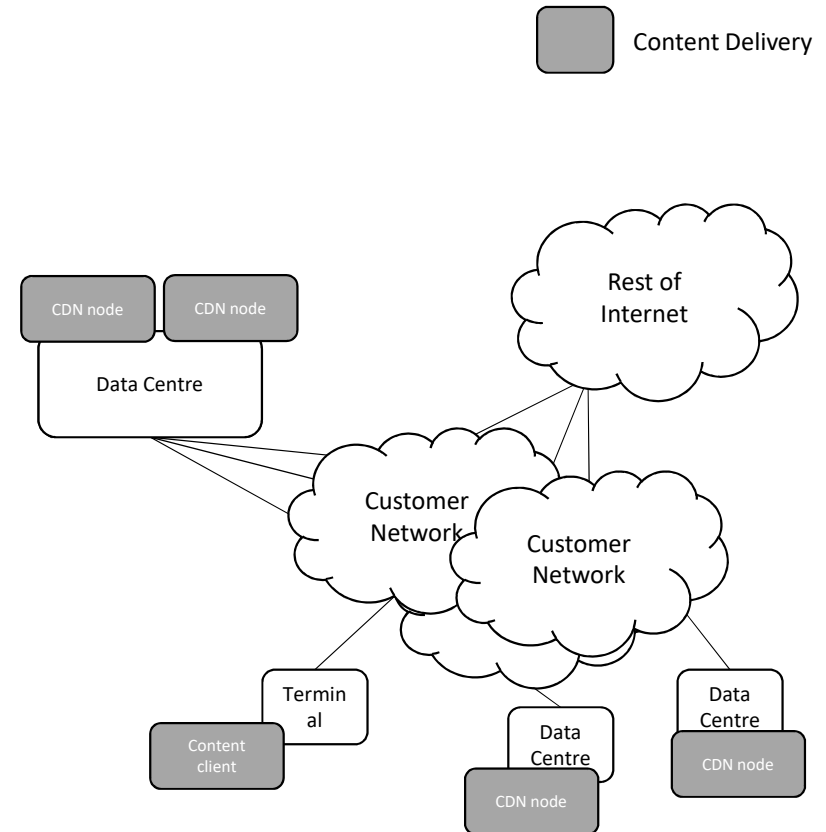Significant increase in media content, both in catalogue size and volume

Examples
(1) CDNs in converged 5G, specifically utilizing 5GLAN [1] improving existing CDN solutions in fixed access

Current Technology Pain Points [2]
- *Efficiency* both in terms of network utilization (back and fronthaul) and CDN server costs
- *Latency*, caused by inefficient path length and DNS redirection to CDN
- *Storage* capacity, particularly for edge caching scenarios

[1] https://www.techsciresearch.com/report/global-content-delivery-network-market/1590.html
[2] see https://arxiv.org/pdf/1803.00876.pdf for an evaluation along those pain points

Content Delivery

CDN node   CDN node
Data Centre

Rest of Internet

Customer Network

Customer Network

Terminal
Content client

Data Centre
CDN node

Data Centre
CDN node

Compute-Fabric-as-a-Service (CFaaS)

Main Idea
Build dynamic, multi-technology data centre connectivity across all access technologies as foundational proposition for other value-add use cases
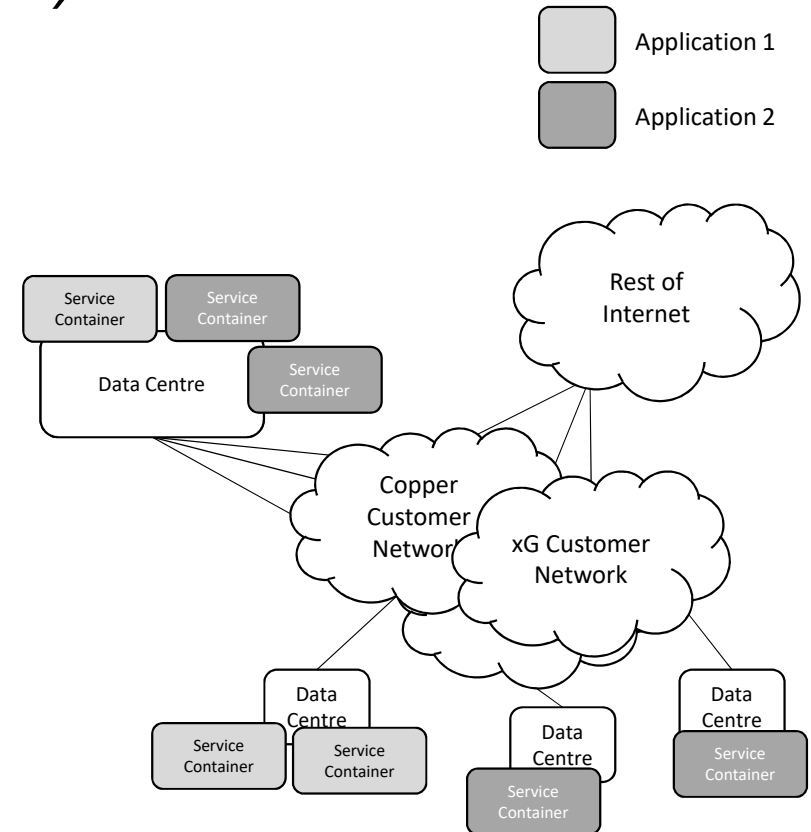
Drivers
application-agnostic infrastructure, utilizing the benefit of available local compute resources (real estate play)

Examples
Almost any vertical, e.g., (1) industrial IoT, (2) gaming (for consumer networks) (3) smart cities

Current Technology Pain Points
- *Dealing with topology changes* to enable service chaining and forwarding across dynamically formed fabric
- *Dynamic addition* of resources, enabling marketplace for local compute resources through dynamic bidding akin to ad bidding on blank webpage

Application 1

Application 2

Service Container

Service Container

Data Centre

Service Container

Rest of Internet

Copper Customer Network

xG Customer Network

Data Centre

Service Container

Service Container

Data Centre

Service Container

Data Centre

Service Container

Requirements

- Covering areas such as
  - > Service routing
  - > Service chaining
  - > Service execution pinning
  - > Service packaging and deployment
  - > Synchronization and consistency
  - > Placement of execution logic
  - > Support for app-specific invocation protocol
  - > Utilization of L2 multicast capabilities
  - > Specification of tenant requirements
  - > Dynamic integration of resources
  - > Optimize interconnection of resources
  - > Accountability of resource usage

Future Plans

- Extend on research challenges, linking more clearly to the requirements and use cases

- Outline high-level solutions, both existing and under research & development