Optimizing network service design aspects in 5G and beyond scenario

PhD Course - 5G, beyond 5G and 6G: the next frontier (final presentation)

Student: Alessandra Fais (alessandra.fais@phd.unipi.it)

PhD Program in Information Engineering
University of Pisa

Outline

- 5G and beyond: use cases and requirements
- Network cloudification
- Improving round trip latency
- Managing virtualized resources
- Network agility and adaptability

5G and beyond

- Whole new set of *compute intensive applications* for
  - virtual/augmented reality (VR/AR), autonomous vehicles, smart cities/buildings, IoT, tactile internet, wireless sensors/actuators networks, ...
- Increasing in data volume, velocity, variety, veracity, value (*5Vs of Big-Data*) + billions of *connected devices*
- Real-time response is a strict requirement for some applications
  - Extremely high data rate, low E2E latency, high energy efficiency, mobility (speed), reliability, ...

Network Cloudification

- **Offload computation** from the UE to
  - Cloud
    - high latency, network congestion, privacy concerns
  - Multi-access Edge Computing (MEC)
    - Computation and storage capabilities close to mobile users
    - Contextual information for optimizing network and QoS
    - Distributed, decentralized, virtualized

- Distribute big-data/latency-critical computations to MEC servers + transfer compute-intensive and delay-tolerant tasks to the cloud

Improving round trip latency

- Transmission delay
  - Depends on the distance
  - **Exploit proximity (local computation and storage at MEC)**
- Processing and storage
  - Depends on the computation response time
  - **Exploit parallelism, HPC, Data Stream Processing (DaSP)**

Idea: apply DaSP to implement real-time applications (e.g., data analysis, anomaly detection, complex event processing in IoT, smart cities, ...) processing streams and deploy them on MEC servers + exploit caching

Managing Virtualized Resources

What is needed?

- Exploit at best the physical server resources
- Support multiple users at the same time while guaranteeing security
- Support live migration of resources and easy deployment

How it can be achieved? → exploiting virtualization!

- Virtual machines VS containers + volumes
  - Impact on scalability and flexibility
  - Different dimensions, complexity
  - Persistent state for containers is maintained in separate volumes
  - Impact on migration is smaller with containers: guarantee service continuity!

Network Agility and Adaptability

- Fastly react to *topology and state variations* of the network
- Efficiently manage user *mobility* (HO) ensuring *E2E service continuity*
- Rapidly accommodate requests for *new services* (concurrently)

How?

- Network service design: cloud native + microservices approaches
- Predict the right amount of resources and the right place
  - Better to estimate the needed resources and add new ones when necessary than overprovisioning (costly)
  - Good allocation of resources in a proactive way
- Proactive and adaptive re-configuration and agility of the network
  - ML/AI → federated learning

AI and its key role in 6G

Joint optimization of 4Cs: communication, control, computation, caching

How to choose the right AI/ML models?

- Centralized approaches worsen data tsunami, hinder scalability, increase latency and exposes user data (privacy concerns)

- Decentralized and distributed approaches
  - *Federated learning* is the most promising approach
  - Collaboration: shared prediction model (and aggregated data)
  - Local data: guaranteed privacy

Conclusions

- Reduce IT response time by exploiting parallelism, HPC, DaSP
  - Optimize both throughput and latency
  - Can be applied both at the edge and in cloud (virtualization!)
  - Integration with content caching is important
- Modular applications → microservices!
  - Right choice of virtualization techniques for each context
  - Proactive allocation and migration strategies → TO DO
  - AI/ML distributed models → must guarantee strict performance requirements and avoid common issues (uplink storm)