5G Technologies and Beyond

Yaru Fu

Assistant Professor
School of Science and Technology (S&T)
Hong Kong Metropolitan University (HKMU)

Apr. 2022

Outlines

- 1. Massive MIMO and mm-Wave
- 2. Wireless Edge Caching
- 3. Wireless Edge Computing
- 4. Distributed Service-aware Network
- 5. Reconfigurable Intelligent Surfaces

- Massive MIMO is an extension of MIMO technology, which involves using hundreds and even thousands of antennas attached to a base station to improve spectral efficiency and throughput.
- Nowadays, most wireless systems are allocated in the band spectrum of 300 MHz to 3 GHz, which is full. The idea behind mm-Wave communications is to take advantage of the unexploited high-frequency mm-wave spectrum, ranging from 3-300 GHz.
    - For security purposes, massive MIMO gives a very oriented beam guide to the legitimate user's location. So, the information leakage is reduced to undesired locations (i.e., Eve) significantly.

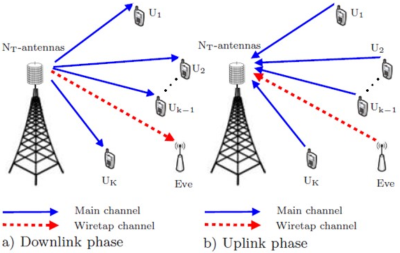
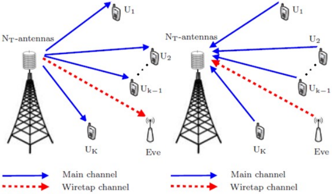


Figure 1: Massive MIMO downlink with K legitimate user nodes, and an eavesdropper.

1. Massive MIMO and mm-Wave

- With CSI, compared with conventional MIMO, massive MIMO is inherently more secure, as the large-scale antenna array equipped at the transmitter (Alice) can accurately focus a narrow and directional information beam on the intended terminal (Bob), such that the received signal power at Bob is several orders of magnitude higher than that at any incoherent passive eavesdropper (Eve).
- Unlike the traditional MIMO, massive MIMO presents the following big challenges:
  - 1) CSI estimation process is a difficult task
  - 2) the channels models are not independent as the distances of antennas are shorter than a half of the wavelength



a) Downlink phase b) Uplink phase

Challenge:

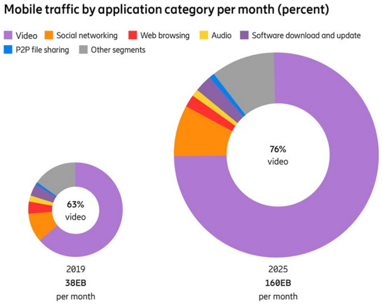


Figure 2: Mobile data traffic prediction by Ericsson

The mobile data traffic in 2025 is 164EB, among which video consumption will represent 76% of the total data traffic in mobile networks.

2. Wireless Edge Caching

Challenges:

- Unsatisfied users' expectation
  - heterogeneous users' demands
  - radio resource competition
- Conventional solutions
  - increase the densification of wireless network infrastructures
    - limited backhaul capacity

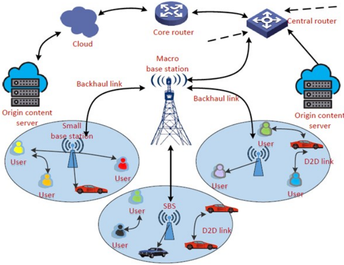


Figure 3: A typical wireless system architecture

- Exploitation of wireless edge caching
  - pre-fetch popular objects and cache them at network edge facilities
    - such as small base stations (SBSs), access points (APs), user devices

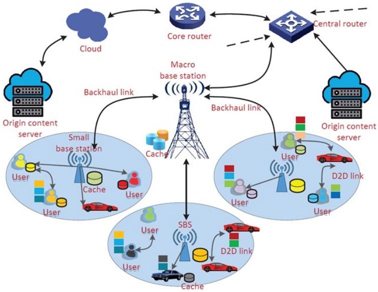


Figure 4: Illustration of wireless caching networks

2. Wireless Edge Caching

Benefits:

1). Reduce users' perceived latency
  - delivering contents from network edge facilities to users experiences *shorter* transmission distances

2). Relieve backhaul link pressure
  - pre-fetching and caching can be done during *off-peak* time

3). Improve users' quality of experience

However, the efficacy of edge caching is only pronounced when the users' preferences are *highly homogeneous* due to

- limited cache capacity at edge facilities
- large number of contents

Next-generation wireless network possess a *higher degree of heterogeneity* due to personality differences.

Recommendation-aware Caching

- Recommendation
  - Recommendation reshape users content consumption patterns
    - the view of videos in YouTube reaches to be 50% with recommendation
    - the percentage increases to be 80% for Netflix
  - A typical recommendation system is to recommend users with the most appealing contents

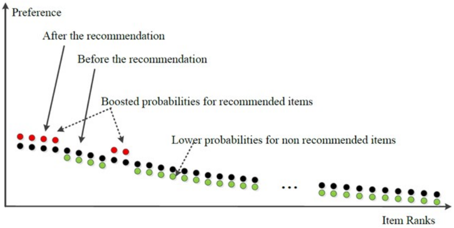


Figure 5: Illustration of recommendation on users' preference distribution

Recommendation-aware Caching

- Recommending each user its top preferred contents is certainly *not optimal* in cache-enabled wireless networks
  - users' preference distributions are highly dispersed and flat
- Nudging users' preference towards better caching performance
  - recommending each user the contents that *not only* align with its own preference *but also* locally cached

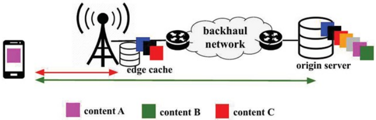


Figure 6: Caching-aware recommendation

- Recommendation affects the content requests of users, which in turn influences the cache placement
  - the two types variables should be optimized jointly. For instance, cache hit ratio maximization, transmission latency minimization.

Reliability-aware Caching

1. Most of the existing works on wireless edge caching assumed that a typical content can *either* be cached *or* not be cached among the edge nodes (ENs)
2. Besides, it was assumed that data corruption at ENs will *not* happen

In practical networks, these two assumptions are *less likely* to be met:

- The content items have large sizes, e.g., high-definition video files
  - The high-definition video files, in general have a large size, caching the entire file in a storage entity will cause a long perceived latency of users, which further increases the abandonment rate
- The cache entities are distributed in a large geographical area and are connected to each other via wireless links. They become unreliable due to the hardware problems, software upgrades or network congestion
  - By Facebook, for a 3000-node cluster, the maximum number of failed nodes per day is > 100

In real systems, contents are divided, coded, and cached among edge nodes.

Repair Consideration: An Example

- A typical content of size B is divided into 4 blocks, i.e., P_1, P_2, P_3 and P_4
- The 4 blocks are encoded into 6 blocks, i.e., P_1, P_2, P_3, P_4, $P_1 + P_3$ and $P_2 + P_4$ by using binary linear code, and stored among 4 ENs

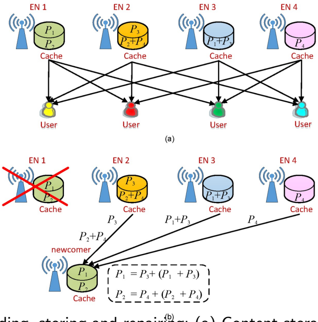


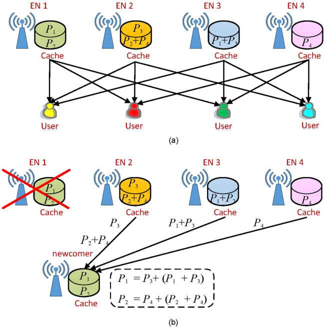
Figure 7: An example of content coding, storing and repairing; (a) Content storage at the edge nodes (ENs) and content retrieval; (b) Content repair

Once a failure of EN occurs, the newcomer needs to download data from surviving nodes to *repair* the lost data such that the system's reliability is ensured, i.e.,

1. any failed node can be repaired successfully
2. any content can be successfully retrieved by accessing ℓ nodes

Repair Consideration: An Example

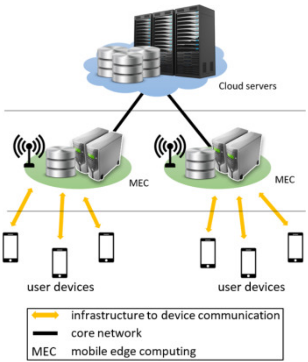
- Storage amount, α_u, $u = 1, 2, \ldots, U$
- Download amount, $\beta(u, v, \mathcal{S}_i)$
  - $\mathcal{S}_i$ depicts *helper set*, where $|\mathcal{S}_i| = d$, $i = 1, 2, \ldots, \binom{U-1}{d}$



$\alpha_1 = \alpha_2 = \frac{B}{2}$, $\alpha_3 = \alpha_4 = \frac{B}{4}$;
$\beta(1, 2, \{2, 3, 4\}) = \frac{B}{2}$, $\beta(1, 3, \{2, 3, 4\}) = \frac{B}{4}$, and $\beta(1, 4, \{2, 3, 4\}) = \frac{B}{4}$

3. Wireless Edge Computing

- Edge computing has been defined as "a part of a distributed computing topology in which information processing is located close to the edge-where things and people produce or consume that information."
- At its basic level, edge computing brings computation and data storage closer to the devices where it's being gathered, rather than relying on a central location that can be thousands of miles away.
- This is done so that data, especially real-time data, does not suffer latency issues that can affect an application's performance. In addition, companies can save money by having the processing done locally, reducing the amount of data that needs to be processed in a centralized or cloud-based location.
- Edge computing was developed due to the exponential growth of IoT devices.



3. Wireless Edge Computing

Computing schemes:

- Local computing
- Edge server computing, which includes three steps, i.e., offloading, computing, and results delivery.
- Partial offloading, e.g., a fraction of α of the task has been processed locally, and $1-\alpha$ has been computed by edge server
  - Joint offloading decision and resource (radio resource, computation resource) allocation; minimize latency, maximize the total computed data or energy efficiency.

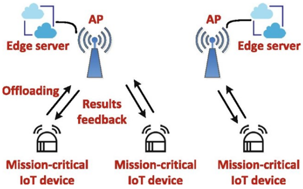


Figure 8: An example of edge computing

Unmanned aerial vehicle (UAV) assisted edge computing

- UAVs have computation tasks
- UAVs act as the mobile edge computing (MEC) servers
- UAVs act as the relay to transmit the computation tasks from devices to MEC servers
  - severe environment, cell edges, railway track surveillance, etc

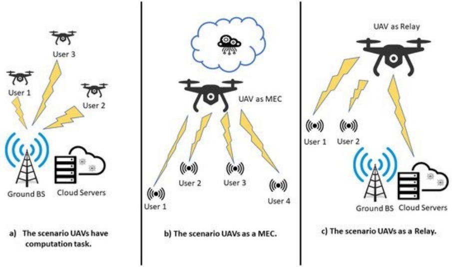


Figure 9: An example of UAV-assisted edge computing system

4. Distributed Service-aware Networks

Challenges:

- As the size of applications or services grows, on one hand, conventional edge computing capability can become sparse very quickly. On the other hand, offloading tasks from user side to the edge servers incurs expensive communication cost, thus cannot satisfy the stringent latency requirements for 6G communications.

To address the above-mentioned challenges:

- European 6G white paper proposed a new paradigm, named distributed service-aware network.
  - To compose each large service as a suite of lightweight and independent microservices (MSs) that perform their respective functions and distribute these functional units among network edges.
  - Once a user requests the service, several edge servers contain different MSs will be commissioned by the central controller to do the computing for this user.
  - Due to the fact that the deployment and resource orchestration for MSs is more flexible and efficient, the intelligent future cellular networks can be realistically expected.

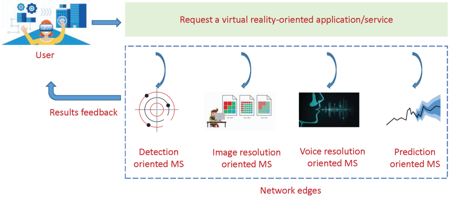


Figure 10: Service v.s. MS

4. Distributed Service-aware Network

Heterogeneous system settings

1. Set up cost, computing cost, and communication cost are different
2. The total capacity of each edge server is different
3. The total computation resource of each edge server is different
4. Different service has different popularity

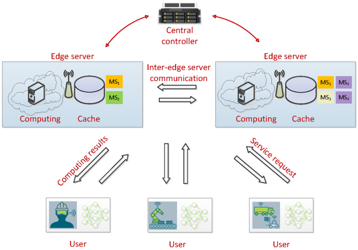
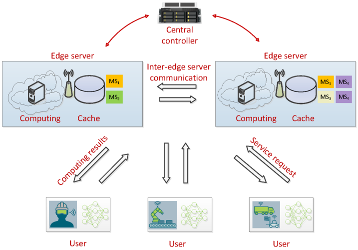


Figure 11: An example of distributed service-aware network

4. Distributed Service-aware Network

The system needs to consider the following issues:

1. MS placement
2. The duplication number of each MS; e.g., each MS can be the component of multiple services
3. The computation schedule
4. Radio resource management



A promising technology for B5G and 6G

1. RIS is a planar meta-surface consisting of a large number of low-cost passive reflecting elements, which can be easily deployed on building facades, roadside billboards, etc
2. By manipulating the phases of incident signals (red lines), RIS has the ability of creating favorable radio propagation environment without additional power consumption

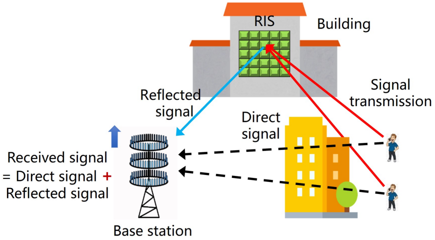
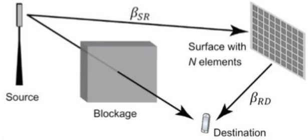


Figure 12: An example of RIS system

5. Reconfigurable Intelligent Surface

$$\text{spectral efficiency} = \log_2(1 + pN^2\beta_{SR}\beta_{RD})$$



Extensive RIS-based research topics:

- MIMO-RIS
- MIMO-RIS-NOMA
- RIS-MEC
- RIS-PLS

Conclusions

B5G/6G is a very broad concept, various techniques, challenges, opportunities exist. Some references are listed below:

[1] "European vision for the 6G network ecosystem," DOI: 10.13140/RG.2.2.19993.95849, Jun. 2021
[2] "6G: The Next Horizon" https://www.huawei.com/en/technology-insights/future-technologies/6g-white-paper
[3] "Imagine the world in 2030 with 6G Flagship White Papers" https://www.6gchannel.com/6g-white-papers/
[4] "DOCOMO 6G White Papers" https://www.docomo.ne.jp/english/corporate/technology/whitepaper_6g/