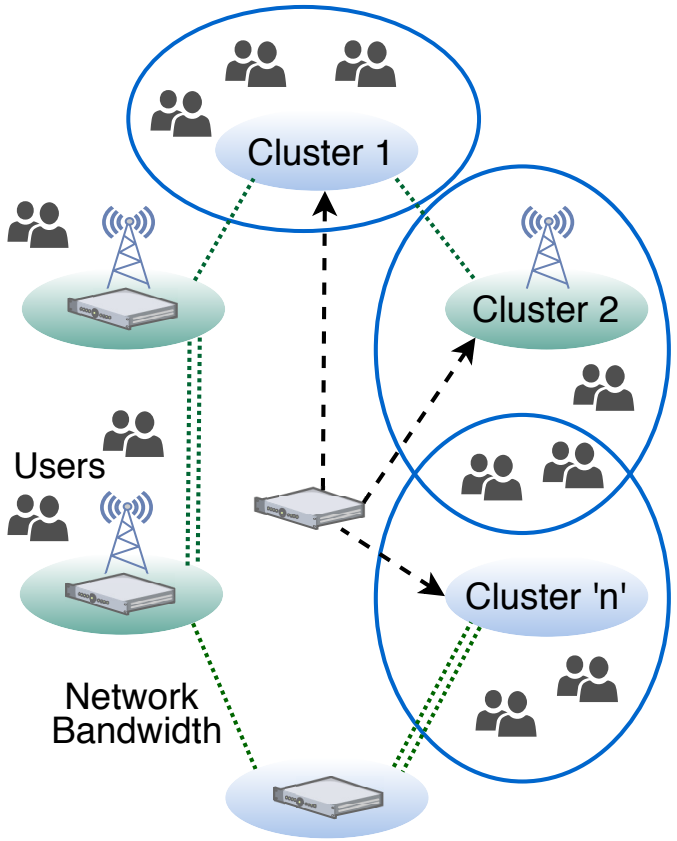
Placing Edge Servers in the Wild!

Placement of Edge Servers

- MEC architectures require edge servers to be co-located with existing base-stations or a new location
- As installing an edge server has a cost:
  1. Electricity
  2. Network
  3. Processing
  4. Maintenance
- Careful analysis is required to find the optimal location and capacity of the edge server which minimizes overall operational cost and network delay from users.

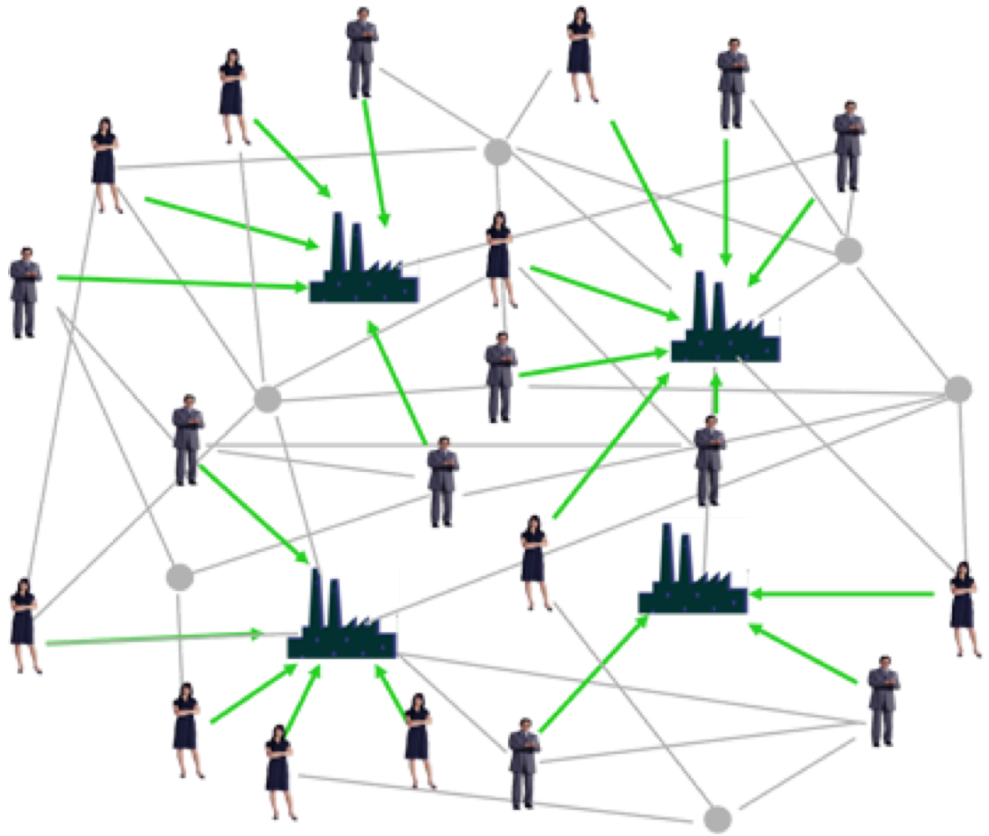
Proposed Approach



- Group existing users in “compute clusters”
- Each cluster center should be the ideal placement for a edge server for that cluster
- Users can belong to multiple clusters.
- Group user clusters until the compute requests do not exceed server capacity
- Formulation resembles Facility Location Problem

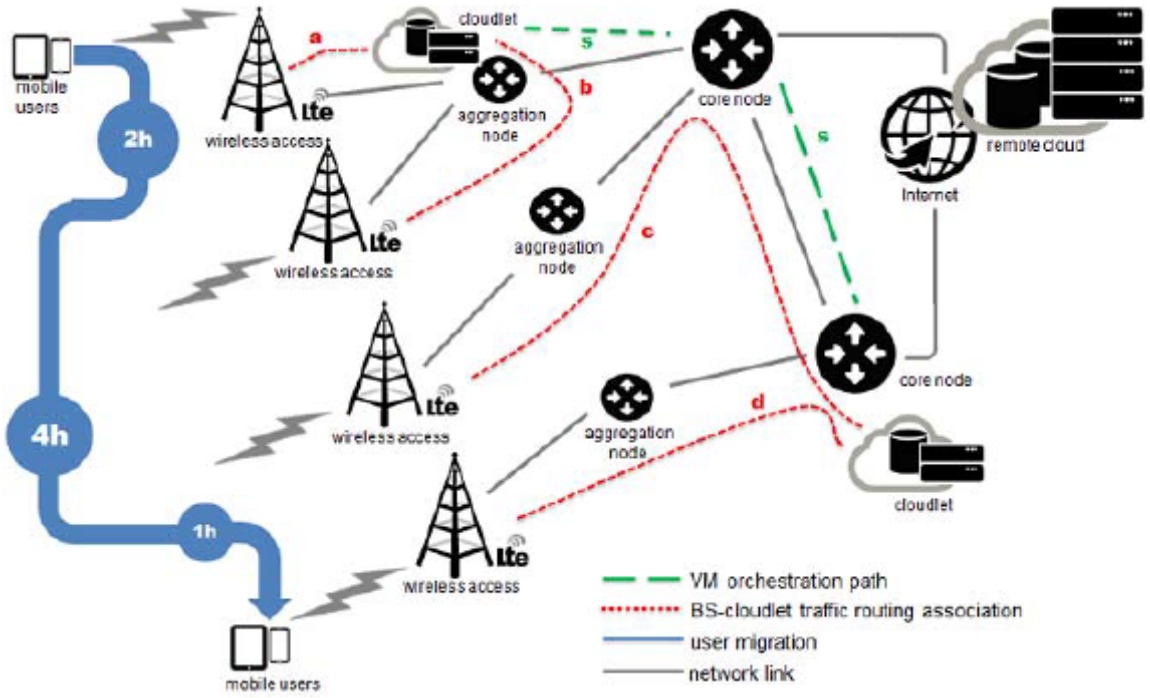
Related Work

Facility Location Problem (FLP)



- Install a facility in a space such that the delay to connect to nearby customers is minimized.
- Each facility can be **capacitated** such that only certain number of customers can be served simultaneously.
- Used significantly in CD edge server placement research
- NP-hard problem

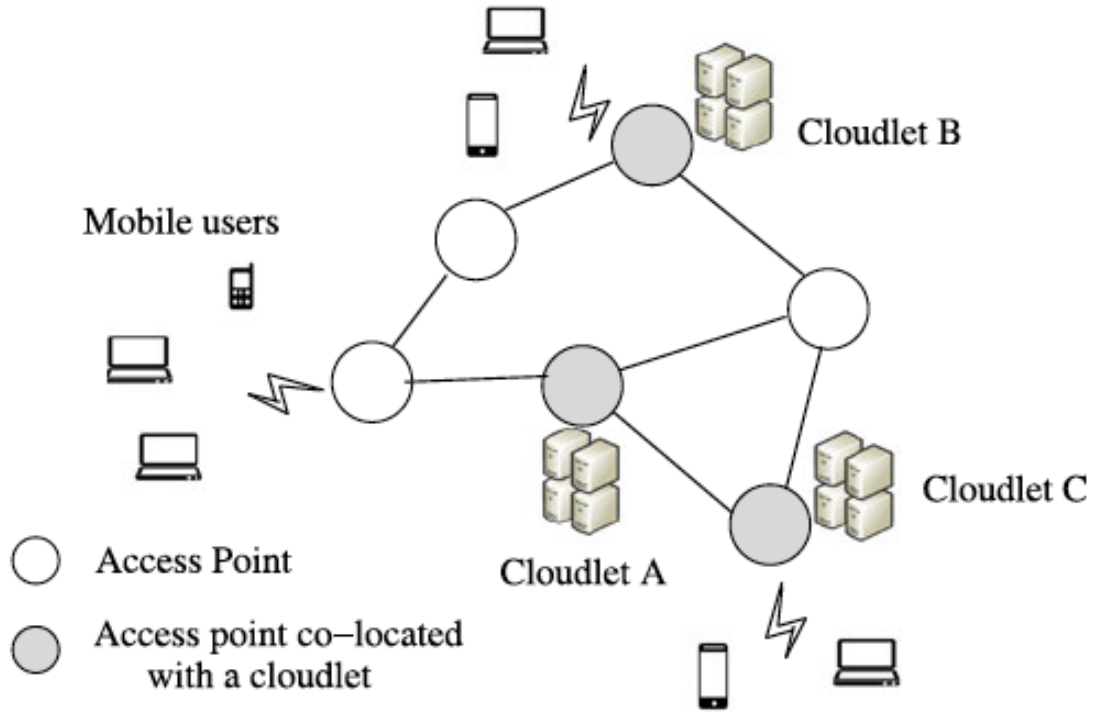
Placing Cloudlets in MEC



- Considers a typical MEC architecture where a cloudlet server can be placed at either aggregator node or core node.
- Models two types of user request pattern: static and dynamic
- Tackles problem:
  1. Cloudlet placement and assignment of Base Stations to cloudlet
  2. VM placement for user request at BS
- Formulation: ILP
- Issues:
  1. The placement node and the graph is known
  2. There is no physical locality and delay constraints between cloudlet and BS

Ceselli, A., Premoli, M., & Secci, S. (2017). Mobile Edge Cloud Network Design Optimization. *IEEE/ACM Transactions on Networking*

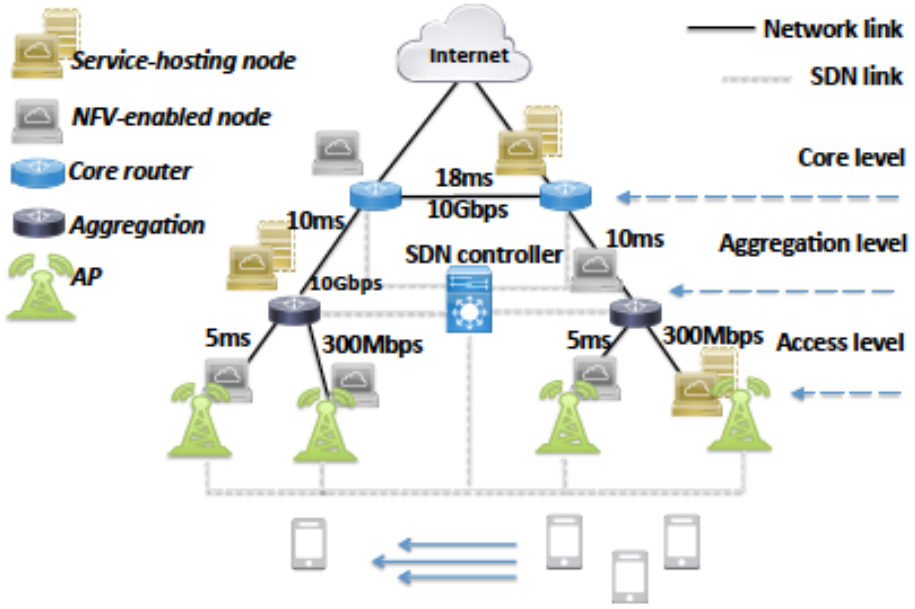
Placing servers on Access Points



- Considers a WMAN network where the Aps are connected with certain network delay.
- Objective to place cloudlet node on AP such that average cloudlet access delay from end-user is minimized.
- Considers cloudlet to be capacitated and user requests having capacity requirements.
- Formulation: K-median ILP
- Issues:
  1. The graph for AP network with delays is known
  2. No user assignment and routing provided
  3. User request pattern is consistent

Xu, Z., Liang, W., Xu, W., Jia, M., & Guo, S. (2016). Efficient Algorithms for Capacitated Cloudlet Placements. *IEEE Transactions on Parallel and Distributed Systems*

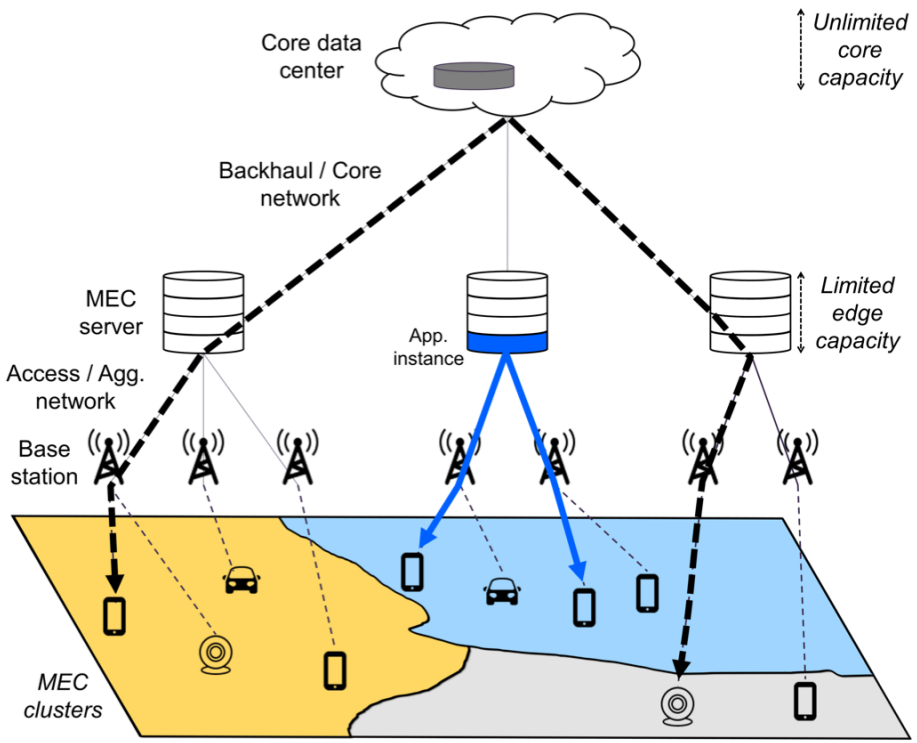
NFV-based Edge VM placement



- Considers an MEC architecture where cloudlets have already been placed and are NFV-enabled.
- Optimizes activating (placing user VM) on cloudlet based on user delay and capacity requirements dynamically.
- Formulation: Knapsack and ILP
- Issues:
  1. The server placement on network has been previously computed
  2. Physical placement problem is not tackled.
  3. User request pattern is pre-known so the assignment is not purely “dynamic”

Yang, B., Chai, W. K., Xu, Z., Katsaros, K. V., & Pavlou, G. (2018). Cost-Efficient NFV-Enabled Mobile Edge-Cloud for Low Latency Mobile Applications. *IEEE Transactions on Network and Service Management*

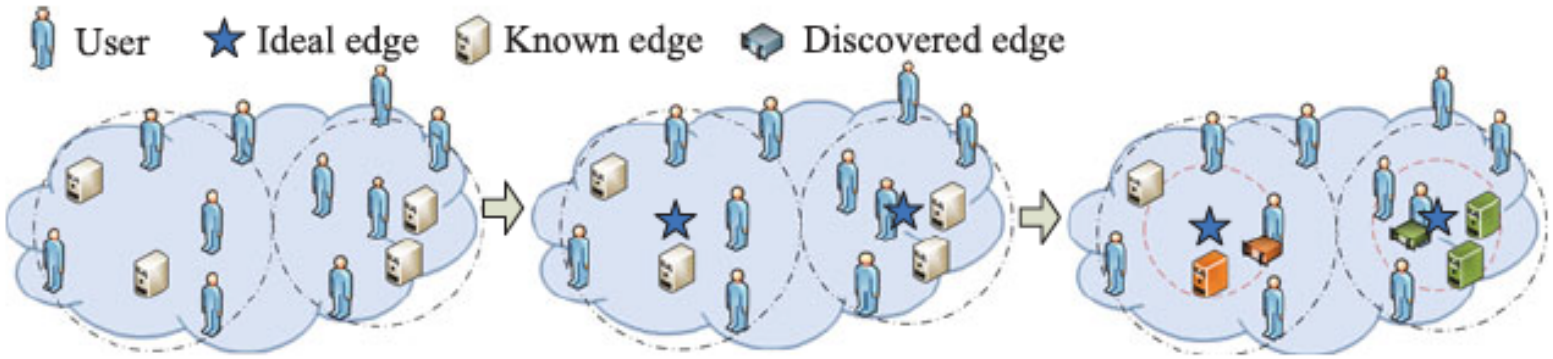
Edge Server placement by user clustering



- Considers and edge network where multiple users are accessing edge servers.
- Assumes that users and servers are co-located in a shared geographical area.
- Optimizes possible edge server placement by dividing users in clusters until overall cluster requests does not exceed edge server capacity.
- Formulation: Graph clustering (Louvain algorithm)
- Issues:
  1. No user mobility.
  2. User request delay within a cluster
  3. One server -> multiple users (not vice versa)
  4. User request placement

Bouet, M., & Conan, V. (2017). Geo-partitioning of MEC Resources. In *Proceedings of the Workshop on Mobile Edge Communications - MECOMM '17*

Finding Ideal Edge Placement

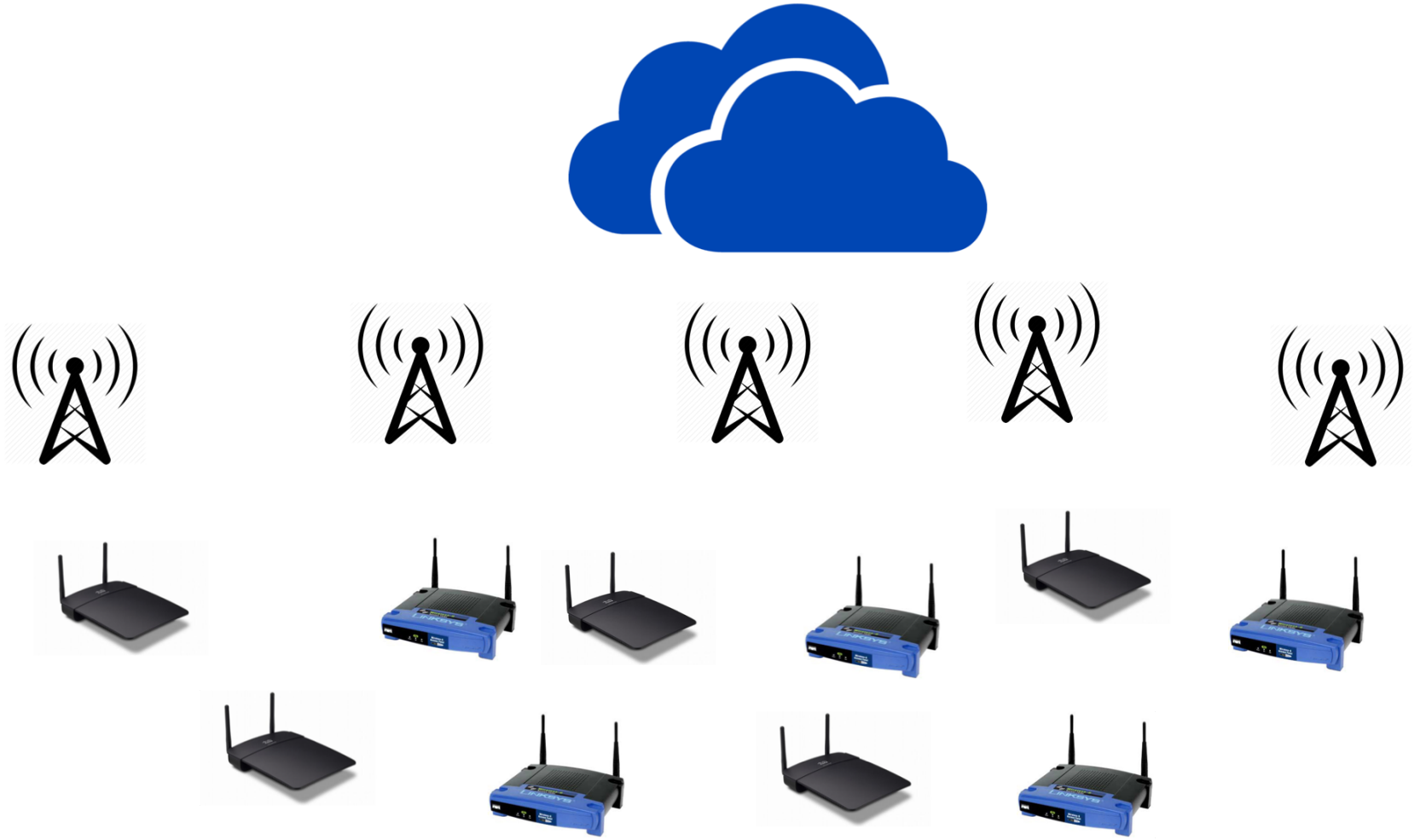


- Assumes a edge network where a server can be placed anywhere.
- Tackles two problems:
  1. Finding ideal placement of server within a user cluster considering network delay (Facility Location Problem).
  2. Mapping ideal location to real network and provisioning according to server capacity constraints (ILP)
- Issues:
  1. Considers CDN servers as Edge so no processing constraints.
  2. User cluster remains consistent

Yin, H., Zhang, X., Liu, H. H., Luo, Y., Tian, C., Zhao, S., & Li, F. (2017). Edge Provisioning with Flexible Server Placement. *IEEE Transactions on Parallel and Distributed Systems*

Revising the idea

Physical Edge server architecture

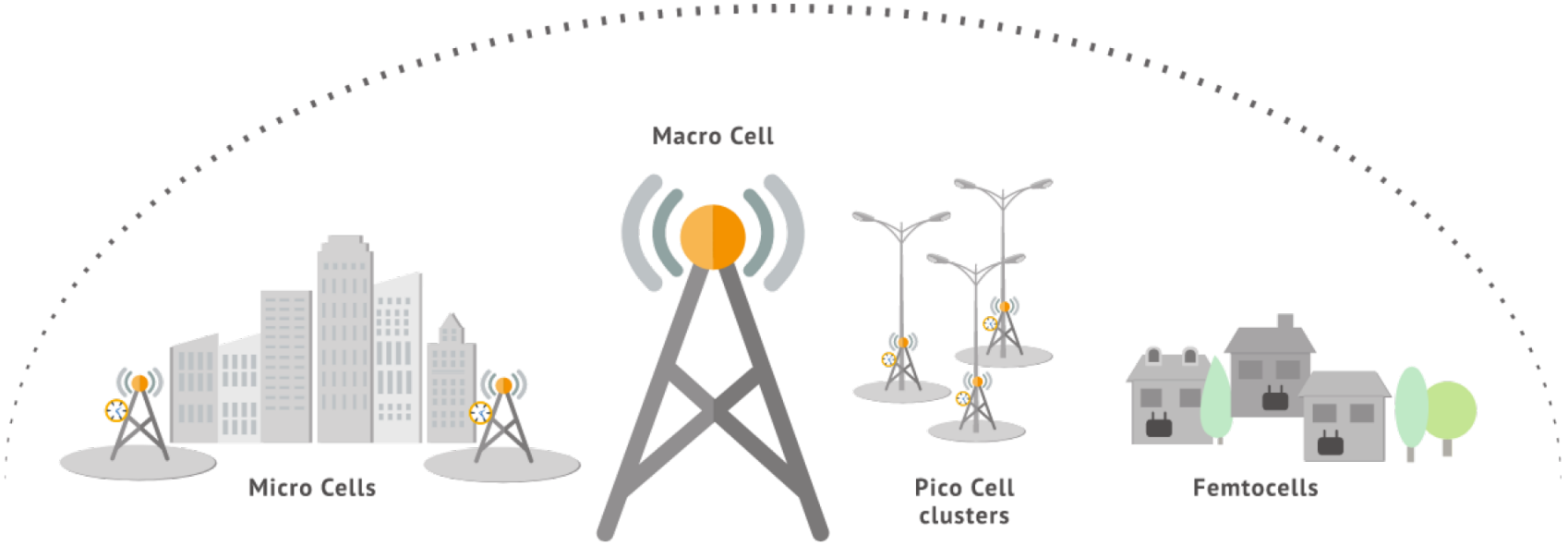


Cloud

Edge-enabled Base Stations

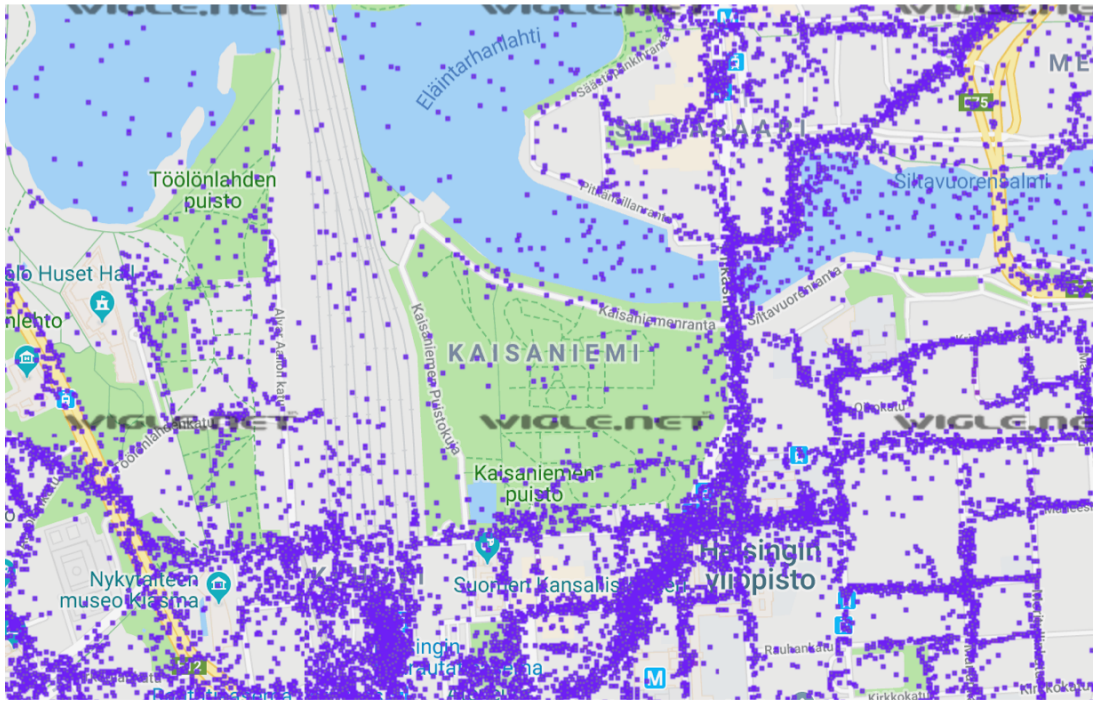
Edge-enabled Access Points

Two-tier Umbrella Architecture



Resembles LTE base station architecture

Server placement on WiFi APs

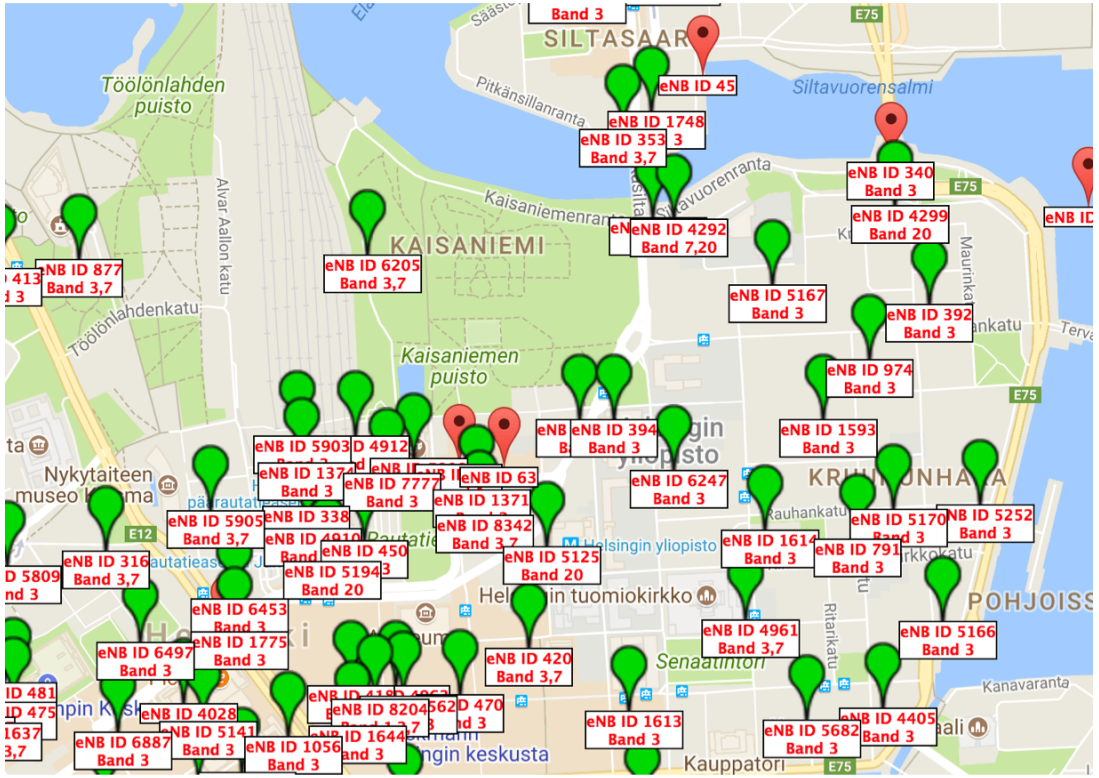


NFV-compliant WiFi APs in a geographical area

1. Get the set of WiFi APs from Wigle
2. Set some APs to have existing edge server with certain capacity
3. Existing techniques can be utilized to get optimal placement.

Issue: We cannot assume that the WiFi based server will have capacity enough to satisfy multiple users.

Server placement on Base Station

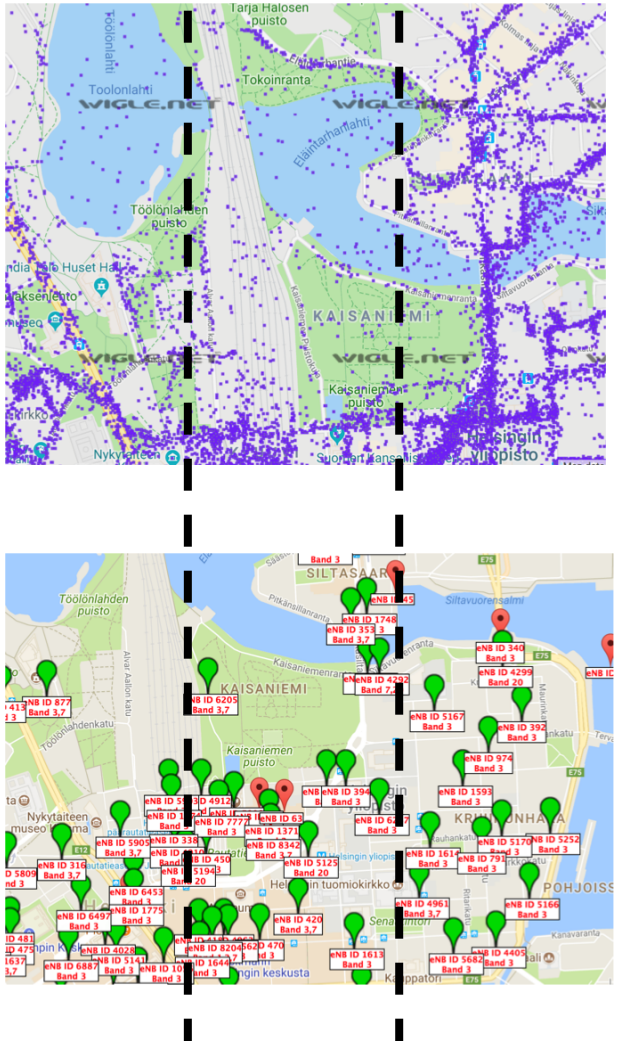


1. Get locations of existing base stations in same geographical location from cellmapper, opencellid etc.
2. Superimpose both base station and WiFi APs in same map locations

Objective

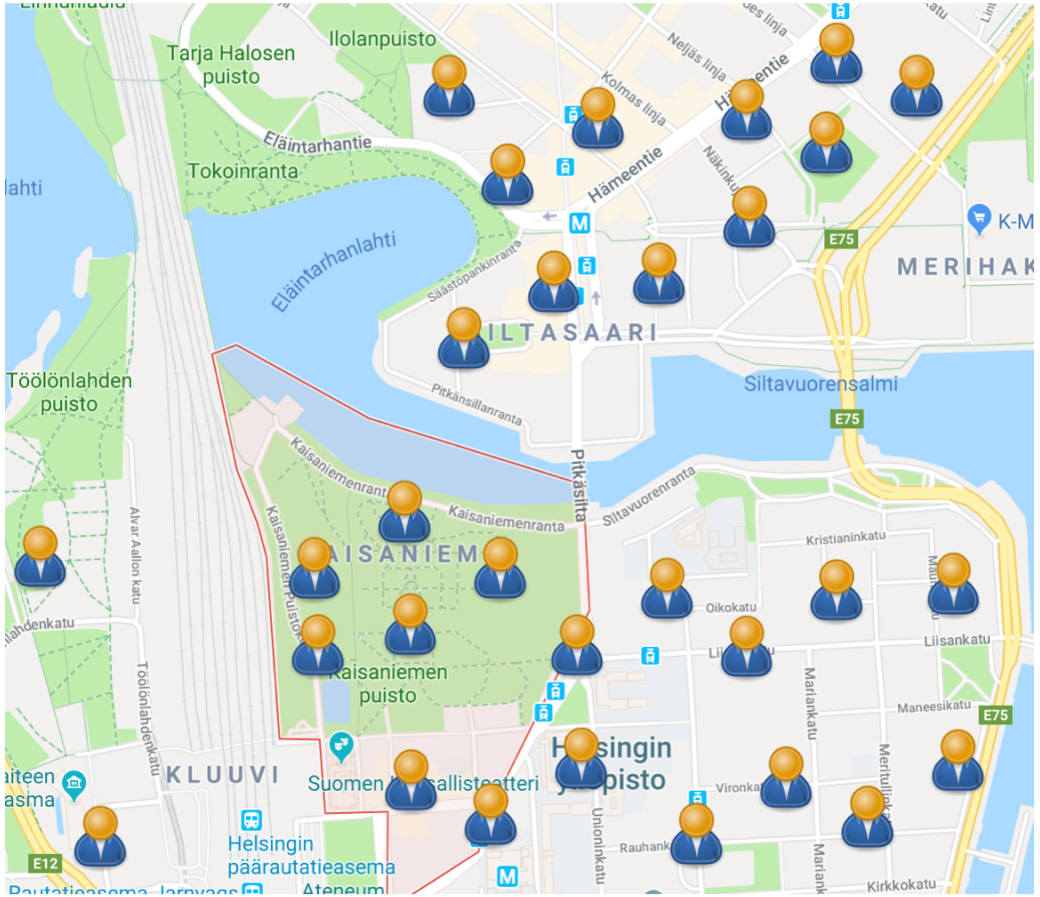
1. Which base stations in the map should host the edge servers and of how much capacity?
   Dependent on:
   a) Server-enabled WiFi APs nearby
   b) Client request density and pattern
   c) Client delay requirements
   d) Capacity of WiFi servers
   e) Cost of server deployment
2. Which user should be assigned to which edge server? WiFi or BS?
   a) Available capacity of server and request by client
   b) Delay requirements imposed?
   c) User mobility? VM migration?

Open Questions to be discussed!



1. How to cluster WiFi APs and base stations?
   a) Geographical locality (graph based grid)
   b) Network delay? (ILP)
   c) User density? (FLP)
   d) Ant-colony?

Open Questions to be discussed!



2. How to design users requests?
   a) Pattern
   b) Capacity requirements
   c) Mobility?

Telecom Italia open database
http://theodi.fbk.eu/openbigdata/#portfolioModal1

Open Questions to be discussed!

3. How to map network delays between WiFi AP network and Base Station network?
4. User allocation to server mathematical formulation
5. Evaluation?