Cooperative Content Caching in 5G Networks with Mobile Edge Computing

Sifat Ferdousi

June 28, 2019

K. Zhang, S. Leng, Y. He, S. Maharjan and Y. Zhang, "Cooperative Content Caching in 5G Networks with Mobile Edge Computing," in *IEEE Wireless Communications*, vol. 25, no. 3, pp. 80-87, June 2018.

- Demand for rich multimedia services has been growing at a tremendous pace – challenging for mobile networks in terms of the need for massive content delivery

- Edge caching - caching and forwarding contents at the edge of networks

- Existing studies treat storage and computing resources separately, and neglect mobility characteristic of both content caching nodes and end users.

- New cooperative edge caching architecture for 5G networks
  - mobile edge computing resources are utilized for enhancing edge caching capability

- Mobility-aware hierarchical caching, where smart vehicles are taken as collaborative caching agents for sharing content cache tasks with base stations

- To further utilize the caching resource of smart vehicles - a new vehicular caching cloud concept
  - vehicle-aided edge caching scheme, where caching and computing resources at wireless network edge are jointly scheduled

- In 5G networks, shrinking of cell sizes and dense deployment of wireless access points open up new opportunities for faster data delivery

- Challenges for growing traffic: centralized nature of mobile network architectures as well as limited transmission capacity entailed by the wireless backhaul links

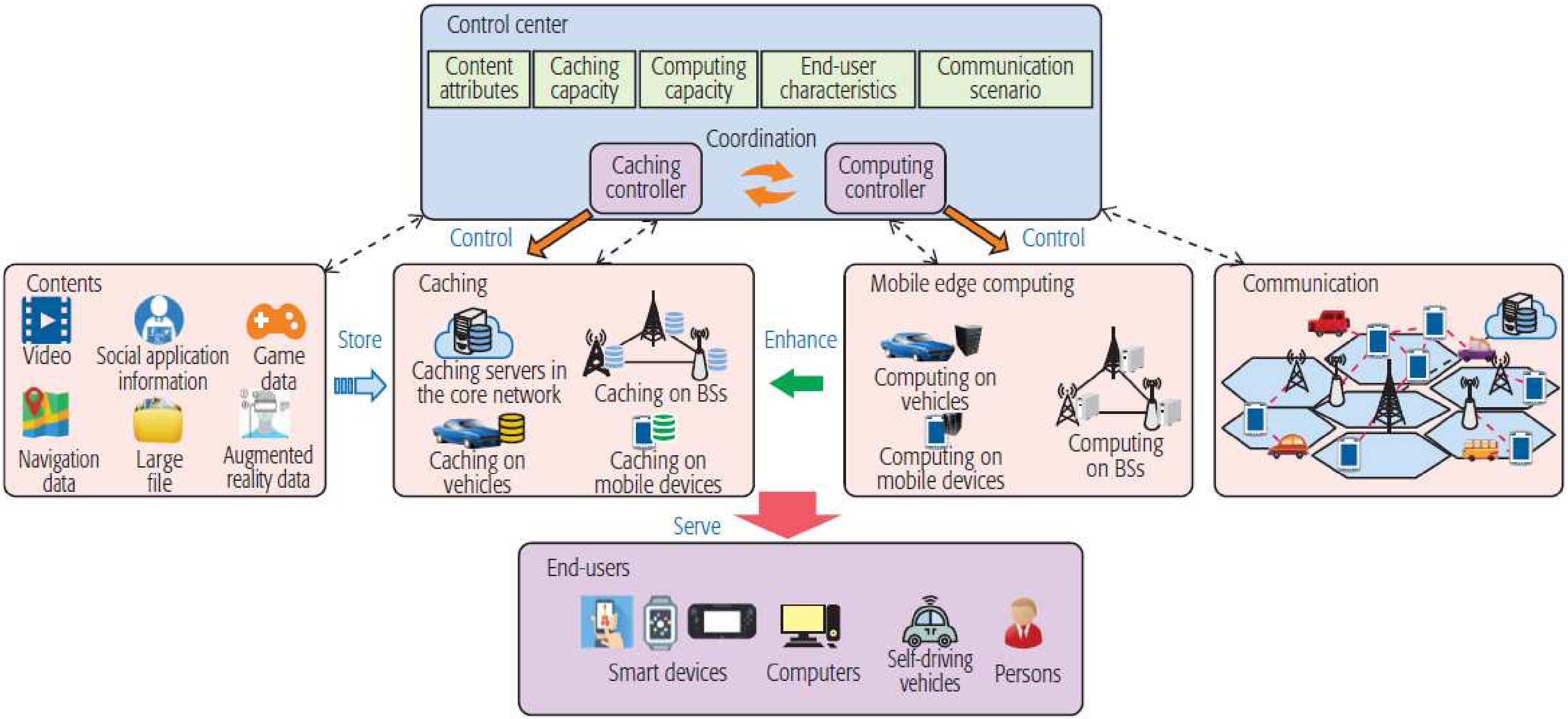
- Although edge caching offers contents to end users in proximity, with ever growing numbers of portable and handheld devices, unpredictable user mobility may heavily affect caching strategies and complicate the content delivery process.
  - *Mobility-aware caching*

- In dense BS deployment 5G networks, during a request for a content of large size, such as a video file, a user with high mobility may pass several small cells.

- Thus, the contents should be optimally cached at edge nodes along the user's path such that it can be fetched by the user when he/she requires it.
- Along with recent advances in wireless communication and IoT, vehicular networks have become an important 5G application.
- Enabled by LTE-V or IEEE 802.11p technologies, a vehicle can communicate with infrastructures, pedestrians, and other vehicles.
- Together with their computing and storage capability, communication-enabled smart vehicles are able to act as moving caching nodes, bringing contents to end users in wide areas.

- New cooperative content caching framework - a hierarchical mobility-aware edge caching scheme that harnesses the synergies between mobile edge computing (MEC), multi-BS caching, and vehicular caching.

- We design mobility-aware edge caching strategies that store popular contents in the BSs passed by mobile end users, consequently minimizing content access delay.
- To further improve the edge caching performance, we exploit content caching and delivery capabilities of moving vehicles.
- Mobility-aware cooperative edge caching scheme that jointly optimizes caching and computing resources of BSs and smart vehicles.


Control center
Content attributes
Caching capacity
Computing capacity
End-user characteristics
Communication scenario
Coordination
Caching controller
Computing controller
Control
Control
Contents
Video
Social application information
Game data
Navigation data
Large file
Augmented reality data
Store
Caching
Caching servers in the core network
Caching on BSs
Caching on vehicles
Caching on mobile devices
Enhance
Mobile edge computing
Computing on vehicles
Computing on mobile devices
Computing on BSs
Communication
Serve
End-users
Smart devices
Computers
Self-driving vehicles
Persons

- Cooperative Caching Architecture shows architecture of proposed caching network. Two types of resources:
  - Caching resources
  - MEC resources

- They serve end users under the management of controllers.
- Caching resources: various types of content-storage-enabled entities.

1. Caching servers located in core network

- Although these servers are far away from end users – they play a vital role as edge nodes always have limited caching capabilities; also popularity of contents is time-varying, so contents cached on edge nodes should be updated adaptively.

- In Internet, contents may be generated from a large amount of providers. If all newly updated contents are obtained directly from providers to edge nodes, high end-to-end latency may be caused by complex interactions between edge nodes and providers, and bandwidth limitation at the content providers.

- Caching servers can help by obtaining and caching new contents according to content popularity. Being intermediate content caching and forwarding devices in core network, these servers can be accessed and utilized easily by the edge nodes.
-
- High bandwidth of core network is helpful to form cooperation of caching servers for sharing their cached contents, which can further improve the caching efficiency.

2. Cache-enabled BSs

- Delivering contents directly to end users, BSs are considered as effective nodes to cache popular contents and reduce duplicate content transmissions from the core network.

- 5G architecture - heterogeneous networks consisting of multiple types of BSs. As various types of BSs have different coverage areas and serve different numbers of users, content caching strategies for each type of BS need to be carefully designed.
  - Compared to microcell BSs, a macrocell BS covers a wider area with more end users. To provide better caching service, contents that meet main requirements of users should be stored on the caching of macrocell BSs
  - Microcell BSs need to follow and cache the particular content demands from local area

- User mobility patterns

- During movement of users, several cells may be passed by. For large content (e.g., video streaming and file sharing), content caching and delivery tasks may be shared by several BSs.

- As the characteristics of the content as well as the moving speed and directions of the users may affect the caching process, how to effectively arrange content segments to cache of the BSs along the way forward is a challenge.

3. Cache-enabled vehicles and mobile devices in caching resources (categorized as mobile caching nodes)

- Smart vehicles and devices have been empowered with caching as well as computing and communication capabilities.
  - Although caching resource of one vehicle or mobile device is limited, accumulative caching power gathered from a group of these mobile nodes is sufficient for storing contents.

- Characteristics of group aggregation and highly dynamic topology of the mobile nodes pose significant challenges on edge caching.
  - The formed cluster of vehicles and mobile devices may be separated due to different directions or different moving speeds of these mobile nodes. Thus, formation and active duration of a mobile node group play an important part in mobile caching utilization.
  - Caching capability together with the communication capacity of various types of mobile nodes are also different.

- MEC, a key technology toward 5G, provides cloud computing capabilities and task offloading service at the edge of mobile networks. Due to the proximity of MEC servers to end users, tasks can be offloaded and accomplished with low latency and high efficiency.

- Similar to the composition of caching resources, in the proposed architecture, MEC resources consist of heterogeneous BSs, smart vehicles, and mobile devices equipped with computation capabilities.

- Although MEC resources seem different from caching resources, they are closely coupled. For instance, using MEC resources on file compressing, the size of a file may be reduced. Thus, some storage space can be saved. From another perspective, the caching capability of nodes is enhanced.

- Another example is augmented reality, where the key elements of the captured video can be extracted from the original data through information processing and computing. As the size of key elements is small, they can be cached and distributed easily. Based on the received key elements, end users may reconstruct the original image or video.

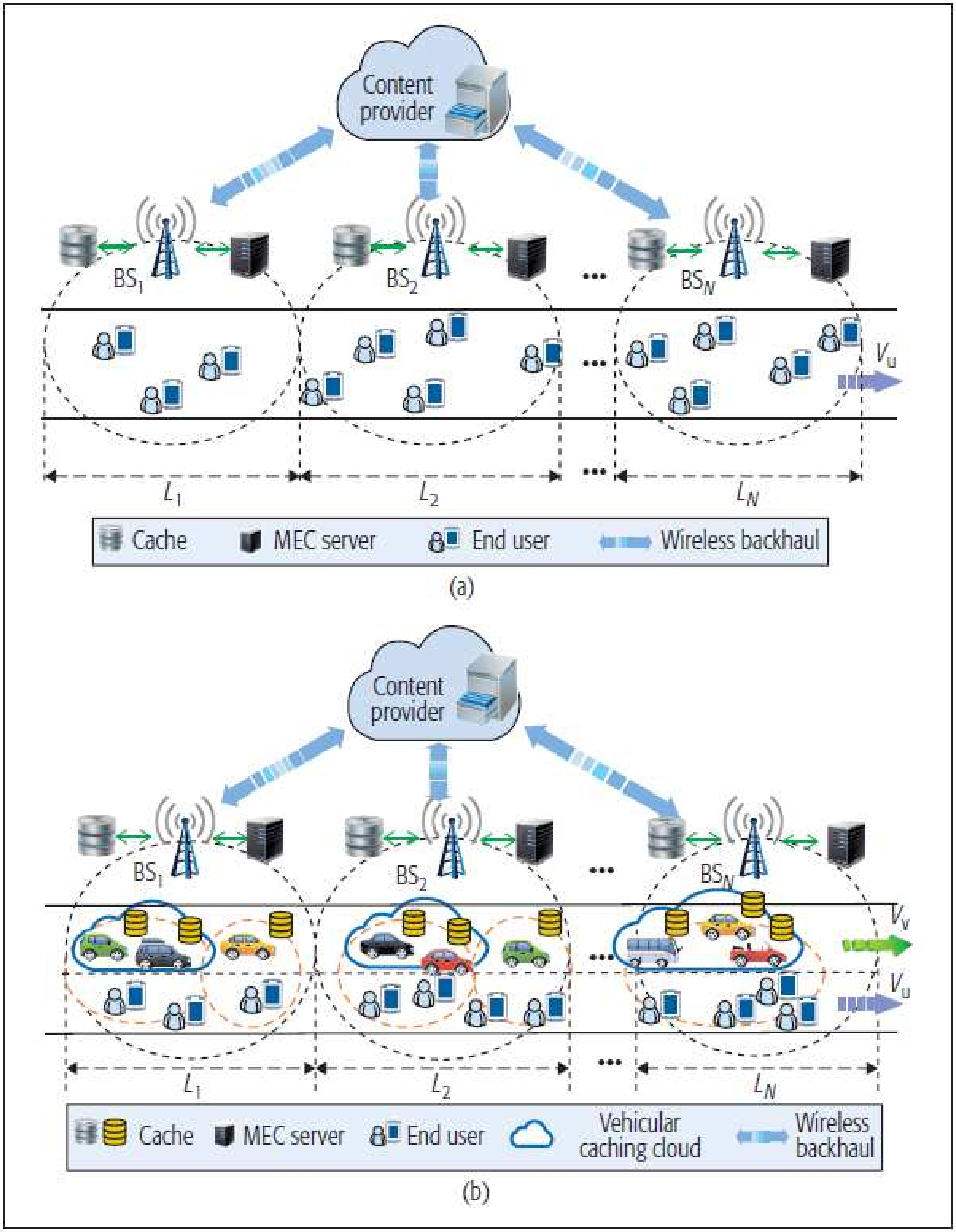
- Machine learning is a feasible approach to address the problem. For instance, a reinforcement learning algorithm can be adopted to adaptively arrange serving capabilities to satisfy the content requirements based on various factors and long-term outcome evaluation.

Mobility-Aware Cooperative Edge Caching Schemes

- Improve end-user experience by reducing their content acquisition time

- Explore both storage and computing capabilities of caching nodes, and raise their storage capabilities by utilizing computing resources
  - considering storage limitations of BSs, to improve content caching capacities, MEC servers can be utilized to reduce size of content files

- Consider mobile characteristics of the vehicles and end users


Content provider
BS$_1$
BS$_2$
BS$_N$
V_u
L_1
L_2
L_N
Cache
MEC server
End user
Wireless backhaul
Content provider
BS$_1$
BS$_2$
BS$_N$
V_v
V_u
L_1
L_2
L_N
Cache
MEC server
End user
Vehicular caching cloud
Wireless backhaul

(a)

(b)

Mobility-Aware Cooperative Edge Caching Schemes

- *Cooperative Edge Caching without Vehicles*

- Set of *N* BSs - each BS is equipped with an MEC server
- Amount of cache resource in one BS - f_b
- Amount of computing resource of an MEC server – c_b

- End users with mobile devices requesting contents are moving along the road at speed V_u
- During movement, end users may pass through several wireless coverage areas of BSs
- Length set of road sections covered by these BSs is $\{L_1, L_2, ..., L_N\}$

- Contents requested by end users are classified into *S* types

Cooperative Edge Caching without Vehicles

- Minimize average latency of content downloading process:

$$\min_{\{x_{i,j},y_{i,j}\}} \sum_{i=1}^{S} \rho i \sum_{j=1}^{J_i^{\max}} \{(L_j / V_u t_b - x_{i,j}) t_c + y_{i,j} t_e + x_{i,j} t_b\} / S$$

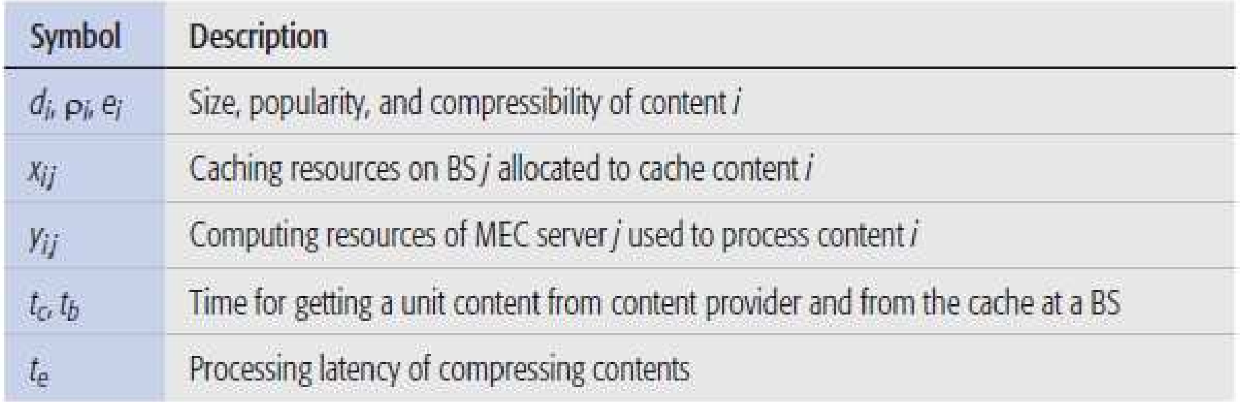
J_i^{max} - index of the farthest BS where type *i* content is cached in

$$\sum_{j=1}^{J_i^{\max}-1} L_j / V_u t_b < d_i \le \sum_{j=1}^{J_i^{\max}} L_j / V_u t_b$$

$$\begin{aligned} \text{s.t.}\quad & \text{C1}: \sum_{i=1}^{S} x_{i,j} / (1 + e_i y_{i,j}) \le f_b, \quad j \in \mathcal{N} \\ & \text{C2}: \sum_{i=1}^{S} y_{i,j} \le c_b, \quad j \in \mathcal{N} \\ & \text{C3}: x_{i,j} \le L_j / V_u t_b, \quad i \in \mathcal{S}, j \in \mathcal{N} \end{aligned} \qquad (1)$$

Allocated caching resource and computing resource are within storage capacity of each BS and computing capacity of each MEC server

Size of content cached in BS j should not exceed maximum amount required by end users



Symbol	Description
d_i, ρ_i, e_i	Size, popularity, and compressibility of content *i*
x_{ij}	Caching resources on BS *j* allocated to cache content *i*
y_{ij}	Computing resources of MEC server *j* used to process content *i*
t_c, t_b	Time for getting a unit content from content provider and from the cache at a BS
t_e	Processing latency of compressing contents

Due to long transmission distance between content provider in core network and end users, $t_c > t_b$

Cooperative Edge Caching without Vehicles

- Each end user randomly chooses one type of contents to download when they arrive at the starting point of the road
  - As users move along the road, they may get part of the content from one BS and other parts from the upcoming ones
  - To provide continuous content delivery service, the contents should be located in the caches of the BSs efficiently

- To solve Eq. 1, use a game theoretic approach to achieve the optimal cooperative caching and computing strategies
  - In this game, players are *S* types of contents
  - Choosing a caching and computing joint strategy, utility of each player is the waiting time to receive the content
  - A Nash equilibrium (NE) of the game is a solution, in which no player can further reduce its waiting time by changing the strategy unilaterally, given the joint strategies of other players
- According to Nash existence theorem, this game possesses at least one pure strategy NE - which is the solution of Eq. 1, in a heuristic manner, where each type of content iteratively updates its joint caching strategy based on strategies of other content types

Mobility-Aware Cooperative Edge Caching Schemes

- *Cooperative Edge Caching Aided by Vehicular Caching Cloud*

- Cache-enabled vehicles can be considered as a new approach to store and spread data - alleviate caching pressure of BSs in 5G networks
  - characteristics of high mobility of vehicles and dynamically changing topology of vehicular networks pose significant challenges

Cooperative Edge Caching Aided by Vehicular Caching Cloud

- Cache-enabled vehicles arrive at the road following a Poisson process

- Let λ be traffic density in terms of vehicles per unit distance - each vehicle has a homogeneous caching resource
- Let f_v be the maximum amount of data that can be stored in the cache of each vehicle

- Speed of vehicles and users are different; during the movement of a user on road section L_j, average number of vehicles passing by the user: $Q = \lfloor \lambda L_j(V_v - V_u)/V_u^2 \rfloor$

- Average amount of data that is delivered by a vehicle to an end user during the passing period: $w = \lambda l/(V_v - V_u)t_v$, where, l and t_v are the length of the road section covered by the vehicle's wireless signal and the time cost for transmitting a data unit from vehicle to end user, respectively

- We assume that the time spent by the user to get a unit data from the vehicle is longer than getting it from the BSs but shorter than getting it from the content provider (i.e., $t_b < t_v < t_c$)

- Considering the caching capacity of the vehicles, active information service capability for one vehicle to an end user can be denoted as $q = \min\{w, f_v\}$

Cooperative Edge Caching Aided by Vehicular Caching Cloud

- To fully exploit caching capability of vehicles and make collaboration between the BSs and the vehicles efficient - vehicular cloud-aided caching scheme.

- The cloud is formed with several cache-enabled vehicles, where the contents are well segmented and stored in these vehicular caches. These cached contents are delivered directly from running vehicles to end users.

- In this way, content receiving latency of end users can be greatly reduced, especially for BSs with poor storage capacity.

Cooperative Edge Caching Aided by Vehicular Caching Cloud

- Heuristic for vehicular caching cloud formation algorithm for content processing and storing:

- *Step 1:* Based on the solution of Eq. 1, for each road section, if there are contents that need to be downloaded from the content provider in the core network, these contents are first divided into blocks with the same size *q*. Then the content blocks may be stored in the cache of the vehicles. One vehicle caches one block.

- *Step 2:* Searching for each road section, for section *j*, $j \in N$, if content type *i* needs processing, compare the time cost of the processing process with that of the data transmission from vehicles to end users. If $t_v < y_{i,j}t_e + t_b/(1 + e_iy_{i,j})$, divide the content into blocks and cache it into vehicles.

- *Step 3:* Caculate the total number of content blocks *Z*. In the traffic, choose {*Z*, *Q*} consecutive arriving vehicles to form the cloud, which store and deliver the content blocks to the end users while passing through the road.

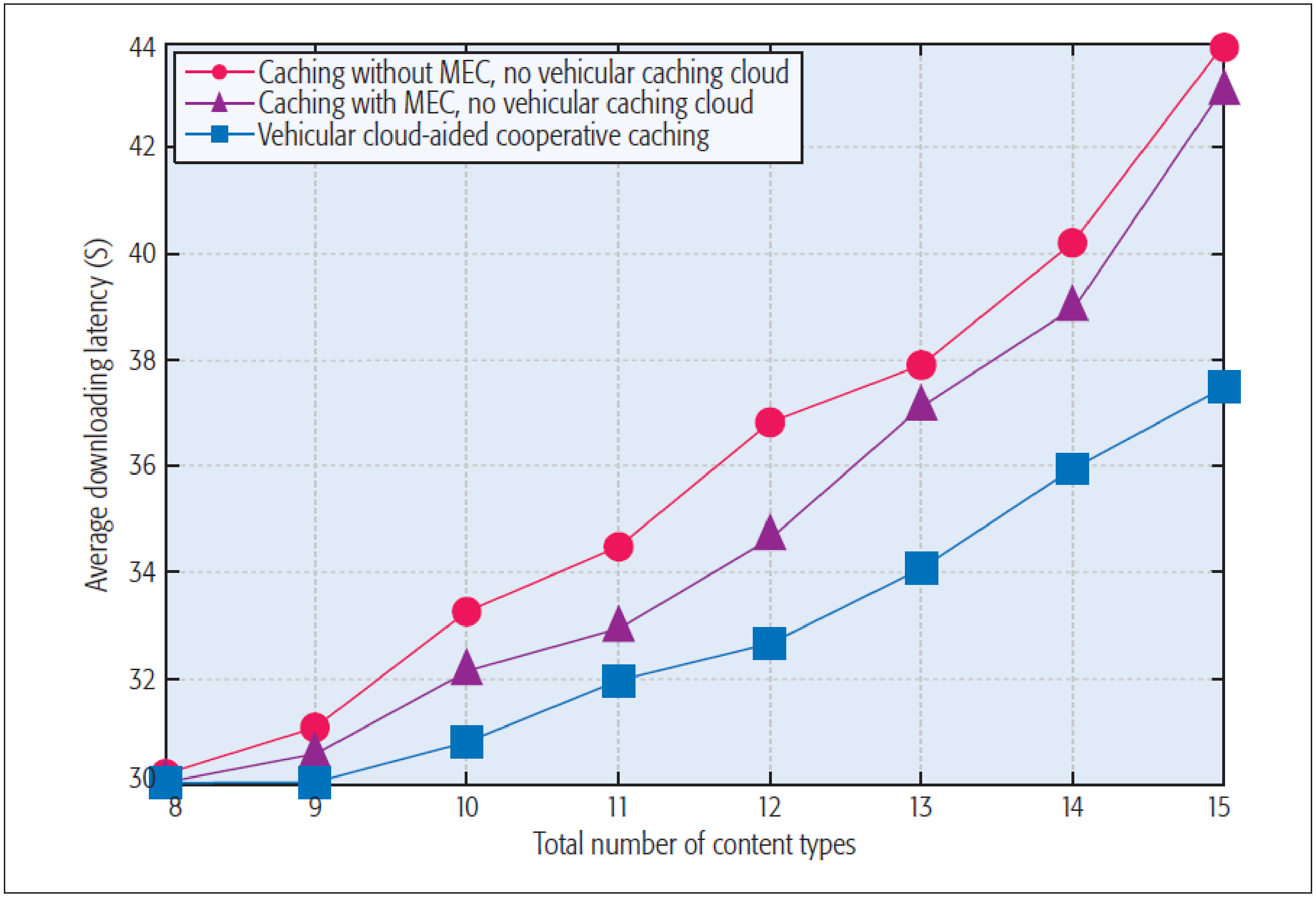
Results

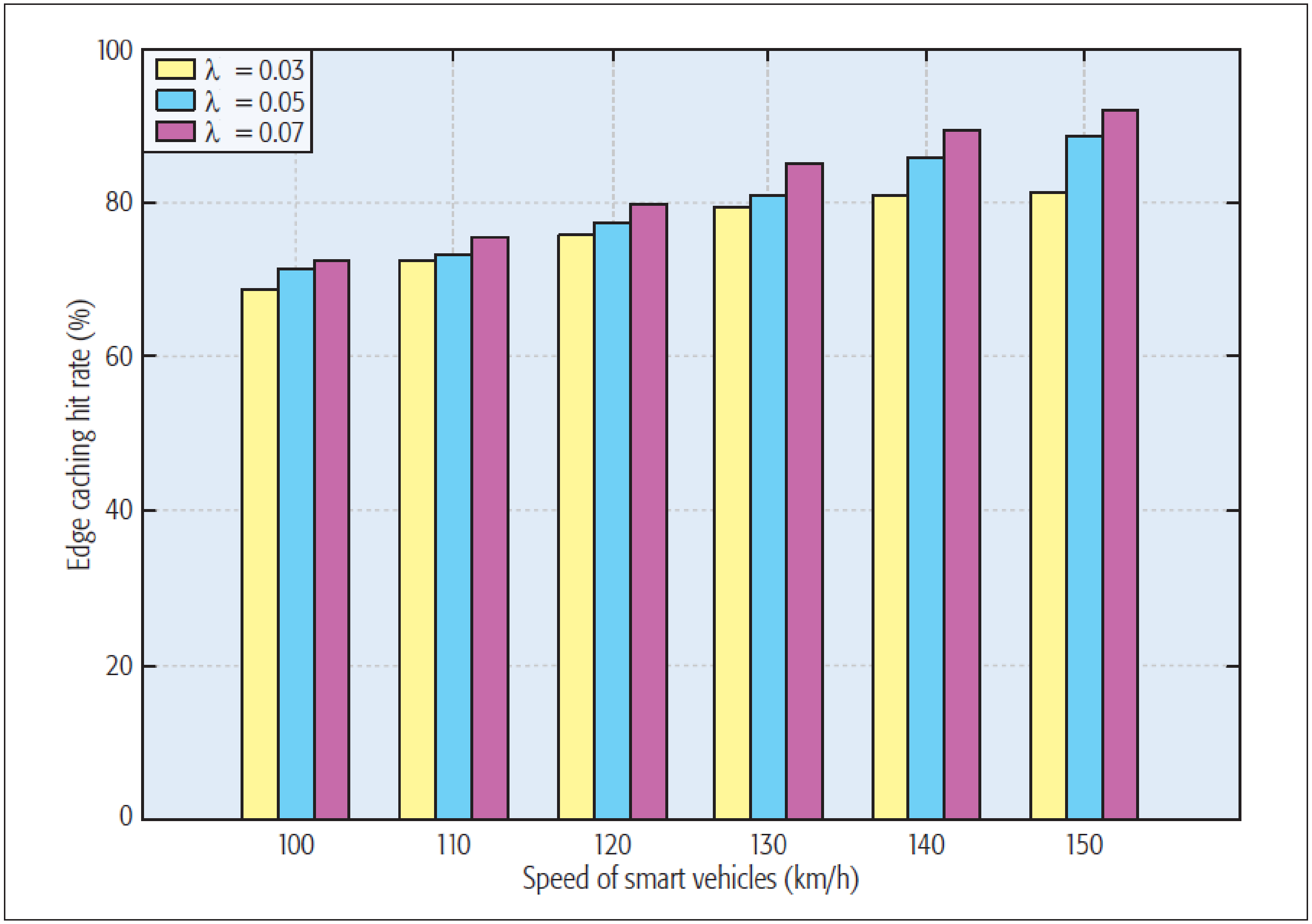
- Consider five BSs located along a unidirectional road

- Caching capacity f_b and MEC capacity c_b of each BS are 1 GB and 50 units, respectively

- End users taking on normal buses move at the speed V_u = 80 km/h, while smart vehicles run at V_v = 120 km/h

- Contents required by end users have large size, which is randomly taken from [500, 1000] MB


Caching without MEC, no vehicular caching cloud
Caching with MEC, no vehicular caching cloud
Vehicular cloud-aided cooperative caching
Average downloading latency (S)
44
42
40
38
36
34
32
30
8
9
10
11
12
13
14
15
Total number of content types


λ = 0.03
λ = 0.05
λ = 0.07
Edge caching hit rate (%)
0
20
40
60
80
100
100
110
120
130
140
150
Speed of smart vehicles (km/h)