Article

An Edge Caching Strategy Based on User Speed and Content Popularity for Mobile Video Streaming

Yingwen Chen [1,†], Hujie Yu [1,*,†] (ID), Bowen Hu [1], Zhimin Duan [1] and Guangtao Xue [2]

1 College of Computer, National University of Defense Technology, Changsha 410000, China;
   csywchen@gmail.com (Y.C.); hbw_14@nudt.edu.cn (B.H.); zmduan0723@gmail.com (Z.D.)
2 School of Electronic Information and Electrical Engineering, Shanghai Jiao Tong University,
   Shanghai 200030, China; xue-gt@cs.sjtu.edu.cn
* Correspondence: yhj258@nudt.edu.cn
† These authors contributed equally to this work.

Abstract: Mobile users' demands to delay-sensitive video streaming media put forward new requirements for mobile networks, such as architecture optimization. Edge caching as a new paradigm is proposed to enhance the quality of service (QoS) for mobile users at the network edge. Due to the limited coverage of edge cache nodes, the frequent handoffs between base stations would aggravate network traffic overhead, resulting in a bad experience of high latency and service interruption when mobile users browse videos. This paper first proposes a three-layer mobile edge network architecture and applied edge caching to video streams to build an efficient caching system. Given the user's mobility and low latency of mobile video streaming, we propose an edge caching strategy based on user speed and content popularity. Horizontally, the user's speed affects the spanning area and the buffer size of the cache on edge; vertically, content popularity determines the priority of cached videos. Experimental results demonstrate that our caching strategy outperforms other schemes in terms of the average delay and the cache hit ratio in mobile video streaming scenes compared with the other three classic caching methods.

Keywords: edge caching; mobile video streaming; user speed perception

1. Introduction

With the rapid development of mobile devices and network technology, we are experiencing an era of explosive mobile data traffic. According to Ericsson's report [1], global mobile video traffic topped 136 EB/month, accounting for 74% of global mobile data by the end of 2024. The demand for large-scale video data information storage and transmission has promoted the development of data centers, especially cloud data centers. The processing method of the traditional cloud computing model is to send data to a remote cloud computing center. After processing and analysis, the data result is sent back to the user terminal [2]. However, according to the Cisco white paper [3], by 2023, the number of devices connected to IP networks will be more than three times the global population, and there will be 8.7 billion handheld or personal mobile devices and 4.4 billion Machine-To-Machine connections.

The rapid growth of networked devices makes the traditional cloud computing model unable to meet the low-latency and high-speed requirements of mobile video traffic. The mobile edge computing (MEC) model takes advantage of storage and computing resources at the edge of the wireless access network to provide services for mobile users, which can make up for the deficiencies of cloud computing [4]. In recent years, edge caching is an effective way to solve the rapid increase in bandwidth demand caused by data-intensive mobile video streaming [5]. Edge caching technology loads data from the data storage into the cache on demand through the mobile edge server, effectively increasing the transmission rate and reducing packet loss and transmission delay. Edge caching

strategy has been a hot research issue, and it is widely used in the industrial internet of things [6], vehicular ad hoc networks (VANETs) [7,8], and mobile edge networks [9–11].

In contrast to conventional live and on-demand video streaming that are consumed using TVs and PCs, mobile video streaming is generally watched by users on mobile devices with wireless connections, i.e., 3G/4G cellular or WiFi [12]. To meet user quality of experience (QoE) for mobile video streaming, edge network services should be stable, reliable, and low-cost [13]. With the rapid development of video streaming media technology, higher requirements are put forward for the smoothness, clarity, and content quality of video playback. Since video streaming is an online service, video latency is the most intuitive factor that affects the quality of user experience. Considering that the video caching service is also online, the cache hit ratio can better reflect whether users can access their favorite videos accurately under the appropriate network conditions and device memory size.

To ensure that mobile users can enjoy video services satisfactorily, we need to pay attention to its mobility characteristics and low delay requirements. As an example shown in Figure 1, in the edge caching scene of mobile video streaming, video content is provided by the content provider or central cloud server in the core network, and then cached in the macro base station (MBS) and the small base station (SBS) through the backhaul link. The SBS closest to the mobile user transmits the cached content to the user equipment via a high-speed link according to the user's request. In the whole process, the MEC server is responsible for caching strategy scheduling, as well as content compression, decompression, encoding, and decoding tasks. As the current map navigation technology is relatively mature, our research on mobile edge cache is based on the premise that the user's moving path is known. Using edge caching strategies to improve mobile video streaming services currently faces the following problems: (1) When the user path is known, arranging the base stations on the route to buffer appropriate video content to minimize the average delay is the key to improving the QoE. (2) When users move, the handoffs between base stations will be caused, which will suddenly increase time delay and cause video service interruption. (3) In the case of limited base station cache capacity, reasonably arranging cache content and updating to improve cache hit ratio is an essential issue for mobile edge cache.
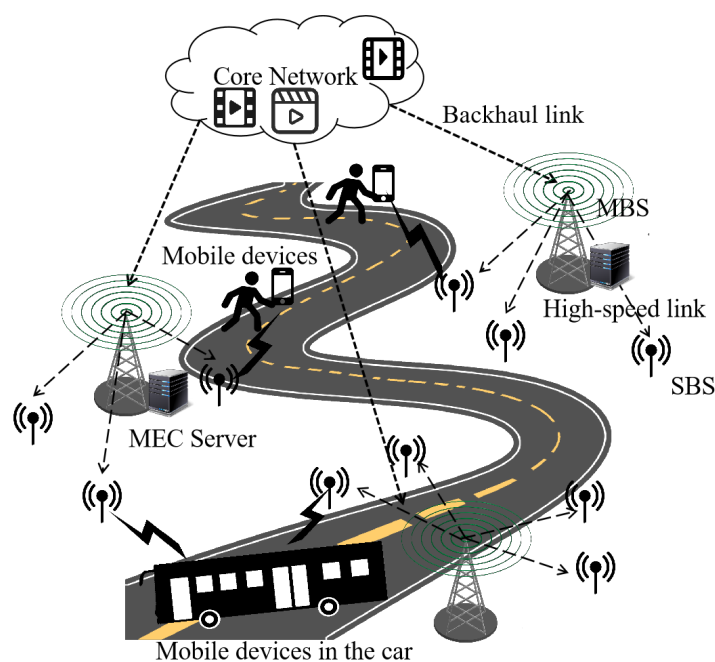

Figure 1. Edge caching scene of mobile video streaming.

Several relevant studies have been conducted to address the above questions. However, such studies are challenging because many different factors are involved, including user mobility pattern, video content characteristics, and mobile edge network characteristics. Previous research usually focused on a single aspect, such as studying the popularity of mobile video content [14,15], user mobility models [16], or mobile network strategies to support mobile video streaming [17]. The limitation of previous studies is that they did not consider the combined effects of user movement, video characteristics, and edge network deployment.

In this paper, we suggest solving the above problems from the perspective of the speed of mobile users and the popularity of video content. From a horizontal point of view, selecting the cached base station on the moving path according to the user's speed and deciding how much video content to cache, which can reduce the high delay and service interruption caused by the handoffs between base stations. From a vertical perspective, according to the popularity of video content, each base station caches high-priority videos, and uses fine-grained and hierarchical storage to improve the cache hit rate. We combined the road traffic model to build a mobile video caching model based on popularity and speed perception.

The main contributions of the paper are summarized as follows:

- First, we propose the hierarchical architecture of the mobile edge cache network, which is divided into three layers: cloud, edge, and user. The edge layer is composed of MBSs and SBSs, and users are distinguished by mobile speed. The hierarchical structure can make full use of the cache space and improve the cache hit ratio.
- Second, we propose an edge caching strategy based on user's speed and popularity for mobile video streaming (ECMSP). When the user path is known, we integrate and optimize the caching mechanism of mobile video streams from the perspectives of video popularity and user movement speed. This strategy can significantly reduce the high delay caused by user movement and service interruption caused by the handoffs between base stations.
- Finally, we implemented simulation experiments to compare our caching strategy with the other three classic caching schemes. Experimental results show that our scheme achieves the best performance in terms of average latency and cache hit rate for mobile video streaming scenes.

This paper is organized as follows: Section 2 introduces the related works. Section 3 presents the architecture and methodology of the proposed edge caching strategy based on user's speed and popularity for mobile video streaming. The experimental evaluation and performance results are presented in Section 4. Section 5 gives some conclusions and discusses potential future work.

2. Related Works

The content distribution technology caches popular video files in intermediate servers and agents, eliminating the need for repeated transmissions from remote servers. This can not only significantly save the transmission resource consumption of the core network but also improve the quality of user experience [18].

Content distribution networks (CDN) have been well investigated in the Internet [19]. The work in [20] characterizes the mechanism by which CDNs serve video content, and the implications for video performance, especially emerging 4K video streaming. Researchers of [21] elaborate on the comparative between HLS (HTTP Live Streaming) and RTMP (Real Time Messaging Protocol) with CDN and also without it. The result shows live video streaming with CDN has better performance than without CDN. However, we cannot simply apply traditional CDN-based content distribution techniques to mobile networks, as legacy CDN-based content distribution mechanism is generally designed for traditional wired communication network architecture [22]. In mobile networks, the resources (e.g., storage, bandwidth, computing capacity) and the position of the deployed servers are constrained. More importantly, the hit rate of cached content items could be rather low in

mobile networks due to the content dynamics, user mobility, and limited number of users in a cell.

An efficient video content caching strategy is of great significance for improving the quality of user experience, and related research are emerging constantly.

Currently, the combination of mobile edge computing and video distribution caching has become a new method of mobile video stream caching, which mainly handles low-latency and intensive computing tasks by transferring cloud computing power and resources to the network edge. Researchers of [23] proposed a distributed caching architecture to bring content closer to the requester to reduce content delivery delays. They designed caching algorithms for multiple operators that cooperate by pooling together their co-located caches, in an effort to aid each other, in order to avoid extensive delays due to downloading content from distant servers. Guan et al. [24] proposed a preference learning-based edge caching strategy (PrefCache). It learns to obtain the user's preference for the video, and, if the video meets the user's needs, it will be placed in the cache. Yang et al. [11] proposed a group-partitioned video caching strategy algorithm (GPC) in vehicular networks. The algorithm first partitions the video requesters and then employs the Lagrange function and Lambert function to solve the cache probability matrix as an optimization variable.

In addition, the key to caching mobile video streams is to pay attention to user mobility. It is the focus of many related works to formulate corresponding caching strategies according to the characteristics of user movement. Su et al. [25] proposed a cross-entropy-based caching scheme in vehicular networks. They analyzed the features of vehicular content requests based on the content access pattern, vehicle velocity, and road traffic density. The work in [26] shows a content prefetching strategy to deal with the dynamic topology in VANETs. The author believes that the pre-caching scheme should be associated with multiple locations and multiple factors. Through machine learning technology, they combine driver preferences with map navigation information to determine the content cache node. Yao et al. [27] chose to use OBU (On Board Unit) as the edge cache node. According to the probability of users' vehicles reaching different hotspots, they proposed a method based on partial matching (PPM) and chose the nodes that stayed in the hotspots for a long time as cache nodes.

Although there have been some works on the cache hit ratio and cache delay in the edge cache of mobile video streams, they often only pay attention to one aspect of mobility or popularity, especially without considering the impact of user movement speed. They also did not consider the impact of the limitation of cache capacity.

3. Architecture and Methodology

In this section, we first give the layered architecture of the mobile edge caching network, and then explain the overall system of the mobile video stream edge caching. Next, we deploy the base station to the road traffic model and perform modeling. Finally, we propose our edge caching strategy for mobile video streams based on user's speed and popularity.

3.1. The Hierarchical Architecture of Mobile Edge Cache Network

Mobile edge caching provides a highly distributed caching environment close to mobile users, which can be used to deploy applications and services, and store and process content [28]. It can effectively place cloud computing and cloud storage on the edge of the network to enable content. Services and applications are accelerated to improve responsiveness from the edge. By caching the data to the local MEC server in advance, users can download the requested content directly from the local cache, thereby reducing redundant transmission, achieving faster service response and higher quality user experience.

Since our caching strategy mainly focuses on the user's speed, we have changed the architecture of the mobile edge caching network, which is divided into three layers, as shown in Figure 2.
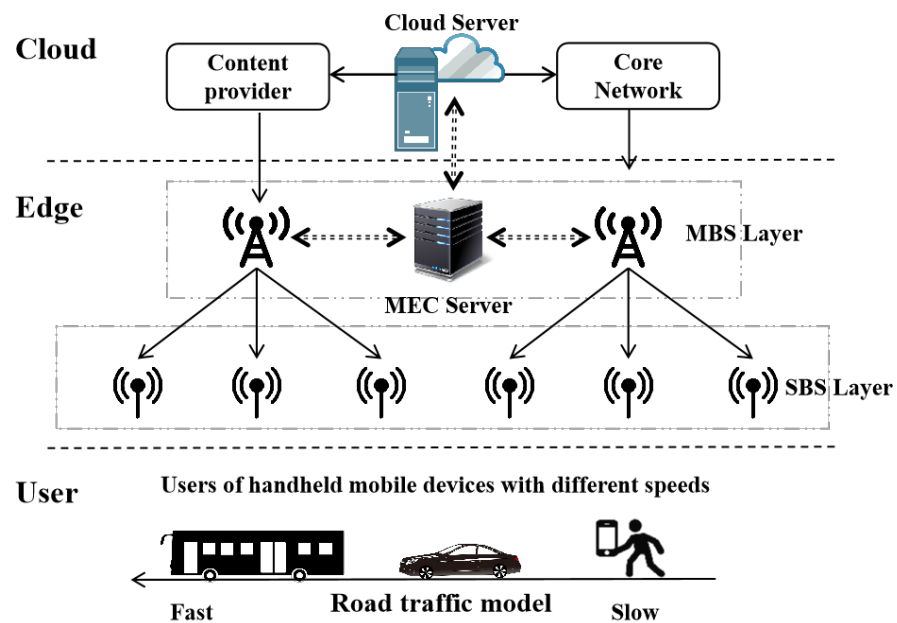
Figure 2. Three-layer structure of a mobile edge cache network.

Cloud Layer: Mainly composed of a central cloud server, and the video content is provided by the content provider of the core network.

Edge Layer: Composed of a MEC server and a base station. The MEC service is responsible for scheduling and coordination, and the base station is responsible for receiving and buffering video content to provide to the user. Base stations are deployed in groups of MBSs and SBSs.

User Layer: Different users are distinguished by different moving speeds, and different speeds have different caching strategies for them. They hold devices that can watch videos and use them to request content from the upper base station.

Based on the above three-layer mobile edge network, we apply mobile edge caching to mobile video streams to form a complete set of efficient caching systems. As shown in Figure 3, the system consists of three parts, video traffic, mobile edge network, and core network. Video content comes with its own bit rate, transmission delay, and popularity attribute to distinguish, which represents the mobile video stream that can be seen on the market at present. The mobile edge network is responsible for managing user information, request queues, network control, and caching and updating strategies. The user model and the video model are combined. That is, each user not only has his own attributes such as moving speed and location, but also makes a request for one of the pieces of video content. The system forms a user video request queue according to the timeline, and transmits the video in the cloud or edge cache to the user device according to the specified location through the task request and response mechanism. Among them, the mobile edge server is responsible for monitoring the operation of the network, and does an excellent job in network congestion control and coordination with the road traffic model. In addition, the cache strategy includes two parts, one is the pre-caching strategy when the cache space is not full, and the other is the cache update strategy after the cache space is full. These three parts perform their duties and coordinate and cooperate to ensure the stable operation of the cache system.
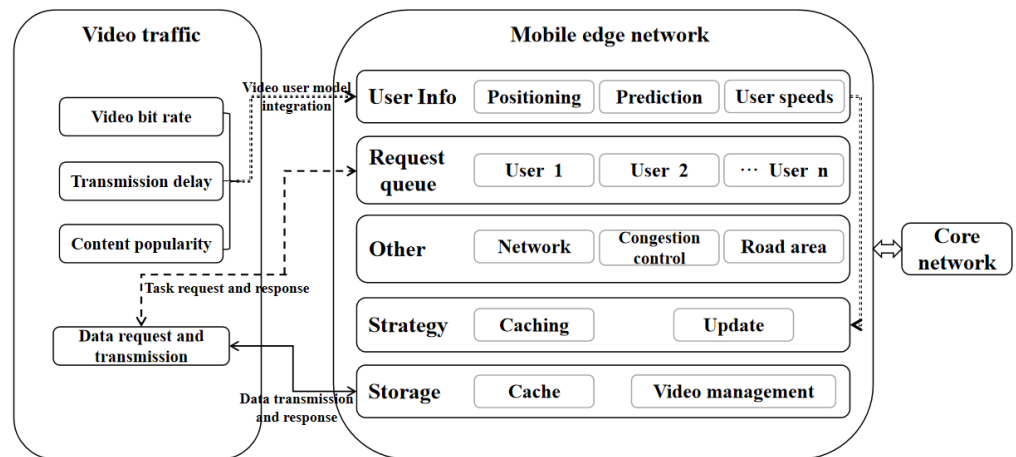
Figure 3. Integral system for edge caching of mobile video streaming.

3.2. Network Model

The deployment method of the base station is shown in Figure 4. Both MBSs and SBSs are deployed on the road grid in a cellular manner. The set of MBSs and SBSs are denoted by $\mathcal{M} = \{1, 2 \ldots m \ldots M\}$ and $\mathcal{N} = \{1, 2 \ldots n \ldots N\}$, respectively. Assuming ignoring the obstruction of buildings, each base station is responsible for a circular area centered on itself, where the coverage radius of MBSs is d_M, and the coverage radius of SBS is d_N. The cache capacity in each base station is limited. The cache capacity of MBSs and SBSs is determined by $\mathcal{R} = \{R_1, R_2 \ldots R_m \ldots R_M\}$ and $\mathcal{S} = \{S_1, S_2 \ldots S_n \ldots S_N\}$, respectively.
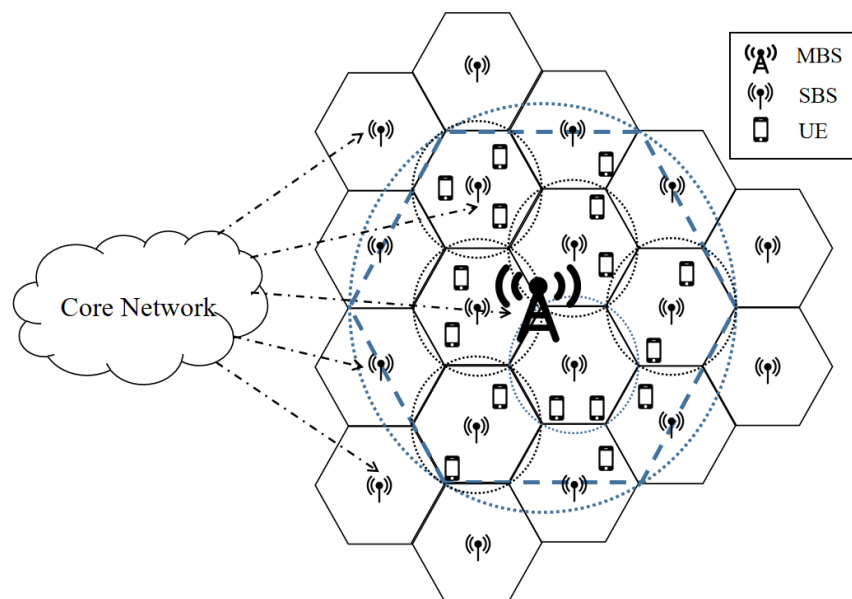

Figure 4. The deployment method of macro base stations (MBSs) and small base stations (SBSs).

Mobile Users: We assume that all users requesting video services are included, conforming to the Poisson Point Process (PPP) distribution. The collection of mobile user equipment (UE) is represented as $\mathcal{U} = \{1, 2 \ldots u \ldots U\}$. Each mobile user is identified with three components, and is presented as $U_u = \{x_u, y_u, v_u\}, u \in \mathcal{U}$, where x_u and y_u are position coordinates of the users. v_u represents the user's moving speed.

Video Content: We resume that each piece of video content has a random size and popularity. The video content collection can be expressed as $\mathcal{F} = \{1, 2 \ldots f \ldots F\}$. Each piece of video content can be expressed as $F_f = \{s_f, p_f, t_f\}, f \in \mathcal{F}$, where s_f and p_f represent the size and popularity of the video content. t_f is the time delay for users to obtain

content. The different popularity of the video means different user request probabilities, expressed as vector $p = \{p_1, \ p_2 \ \ldots p_u\}$:

$$p_j = \frac{1}{\left(\sum_{j=1}^{U} 1/j^{\gamma}\right) j^{\gamma}} \tag{1}$$

where γ is the Zipf index [29].

The transmission distance of the video content can be calculated from the real-time position coordinates of MBSs, SBSs, and mobile users. Among them, the maximum transmission distances from MBSs to SBSs and SBSs to UEs are denoted by d_{MS} and d_{SU}, respectively. The spectrum efficiency performances of MBS m, SBS n, and UE U_u is given by means of

$$r_{m,n} = \log_2\left(1 + \frac{P_{m,n} d_{MS}^{-\alpha_c}}{P_{n,u} d_{MS}^{-\alpha_{cv}} + N_0}\right) \tag{2}$$

$$r_{n,u} = \log_2\left(1 + \frac{P_{n,u} d_{SU}^{-\alpha_v}}{P_{m,n} d_{SU}^{-\alpha_{cv}} + N_0}\right) \tag{3}$$

where $P_{m,n}$ and $P_{n,u}$ are the transmission power. Path loss index corresponding to interference link is expressed as: α_c, α_v, α_{cv}. N_0 is the unilateral power spectral density of additive Gaussian white noise.

3.3. Video Caching Model

3.3.1. Cache Hit Ratio

The cache hit ratio is the most intuitive indicator of the effectiveness of the cache strategy. When a UE requests a video segment in $\mathcal{F}$, we can display cache hit probabilities as follows.

Based on the segmented transmission technology of video streaming media [30], the video content is cached in block grouping. Assume that, in the i-th prevalent sequence, there are video contents of diverse sizes in the $G_i(i \in [1, N])$ group, while the average video size of group $g(g \in [1, G_i])$ is S_{ig}. The ratio of video clips in group g to aggregate files is indicated by s_{ig}. Furthermore, given a cache memory of S_n and a request segment of F_i, the minimum cache hit probability of UEs can be denoted by $S_n p_i c_{nip} s_{ig}$, where the caching strategy for all UEs can be expressed as $C = \left[c_{nip}\right]_{N \times U \times [G_1, G_2, \cdots G_u]^T}$. Hence, the cache hit probability of a video requested to a SBS is

$$h_{hit} = \sum_{n=1}^{N} S_n \sum_{i=1}^{U} p_i \sum_{g=1}^{G_i} s_{ig} c_{nip} \tag{4}$$

3.3.2. Average Delay

The average delay is the most important indicator that affects the quality of user experience. After obtaining the video content F_f, the MBS m sends content f to the SBS n. The delay of acquiring content F_{fp} from m to n can be obtained by

$$del_{m,n} = \frac{s_f}{bw_{m,n}} \cdot d_{m,n} \tag{5}$$

where the transmission distance from m to n is demoted by $d_{m,n}$, and the average bandwidth during the route from m to n is denoted by $bw_{m,n}$. The possibility for sending the copy of video content f can be expressed as $req_{n,f}$, which is given by

$$req_{n,f} = \begin{cases} 1, & \text{If there is a user request} \\ 0, & \text{Other} \end{cases} \tag{6}$$

Here, $req_{n,f} = 1$ means that SBS n has completely cached video content f. The transmission of video content from the SBS to the UE requires a similar process. To sum up, the average delay of video caching is given by

$$del_{\text{total}} = \frac{1}{U} \sum_{n=1}^{U} req_{n,f} \cdot \left(del_{m,n} + \sum_{\varepsilon=1}^{N+1} del_{n,i}^{\varepsilon} \right) \tag{7}$$

3.4. Video Caching Strategy

3.4.1. Problem Derivation

Our research work aims to maximize the hit rate of mobile video streaming cache and minimize the average transmission delay for mobile users to obtain the video. In this section, we propose the problem of maximizing video cache hit ratio and minimizing transmission delay, and present a mobile video cache strategy algorithm based on speed and popularity to obtain optimal variable c_{nip}.

The joint caching hit ratio maximization and caching delay minimization problem is formulated by

$$\begin{aligned}
\mathcal{P} : \max_C \; & p_{\text{total}} \\
= \; & p_{\text{hit}} - p_{\text{delay}} \\
= \; & \sum_{n=1}^{N} S_n \sum_{i=1}^{U} p_i \sum_{g=1}^{G_i} s_{ig} c_{nip} - \frac{1}{U} \sum_{n=1}^{U} req_{n,f} \cdot \left(del_{m,n} + \sum_{\varepsilon=1}^{N+1} del_{n,i}^{\varepsilon} \right) \\
\text{s. } t. \; & \sum_{i=1}^{U} \sum_{g=1}^{G_i} c_{nip} \cdot S_{ig} \leq S_n, 0 \leq c_{nip} \leq 1, n \in [1, N]
\end{aligned} \tag{8}$$

To find the solution to problem $\mathcal{P}$, we first give the following theorem:

Theorem 1. *Under the terms of $n \in [1, N]$, $i \in [1, U]$, $g \in [1, G_i]$, revenue function p_{total} has a maximum cache probability c_{nip} for UE, i.e., $\frac{\partial^2 p_{total}}{\partial c_{nip}^2} < 0$.*

Then, we apply the relaxation variable η_n to find the optimal result by the Karush–Kuhn–Tucher (KKT) condition. With η_n, we can get the Lagrange function as follows:

$$\begin{aligned}
\mathcal{L}(C, \eta) = \; & \sum_{n=1}^{N} S_n \sum_{i=1}^{U} p_i \sum_{g=1}^{G_i} s_{ig} c_{nip} - \\
& \frac{1}{U} \sum_{n=1}^{U} req_{n,f} \cdot \left(del_{m,n} + \sum_{\varepsilon=1}^{N+1} del_{n,i}^{\varepsilon} \right) + \\
& \sum_{k=1}^{N} \eta_k \left(\sum_{i=1}^{U} \sum_{g=1}^{G_i} c_{nip} \cdot S_{ig} - S_k \right)
\end{aligned} \tag{9}$$

In addition, a KKT-based solution to the above optimization problem must satisfy the following conditions:

$$\begin{cases}
\frac{\delta \mathcal{L}(C,\eta)}{\delta c_{nip}} = 0 \\
\eta_k \geq 0, k \in [1, N] \\
\sum_{k=1}^{N} \mu_k \left(\sum_{i=1}^{U} \sum_{g=1}^{G_i} c_{nip} \cdot S_{ig} - S_k \right) = 0
\end{cases} \tag{10}$$

where the first equation represents the necessary condition for obtaining the extreme value. The second denotes the coefficient constraint. The last equation means that the file size must not exceed the RAM size of users. The optimized model above can be evaluated by means of conditional solution $\frac{\delta \mathcal{L}(A,\eta)}{\delta c_{nip}} = 0$.

3.4.2. Algorithm Design

In this part, we first explain the caching mechanism of our proposed caching strategy, and then propose the ECMSP algorithm to maximize the profit function by setting and solving the caching probability matrix.

When a user moves from area A to area C as shown in Figure 5, the duration of the video requires the user to span the area covered by three base stations. Calculate the user's stay time in each area based on user's average speed, and then calculate the size of the video content that needs to be cached in each area based on the transmission rate. In the area at the junction of base stations, additional video content needs to be cached due to speed calculation errors and the handoffs between base stations. When there are multiple types of video content that need to be cached, the content with higher priority is selected according to the popularity of the video.
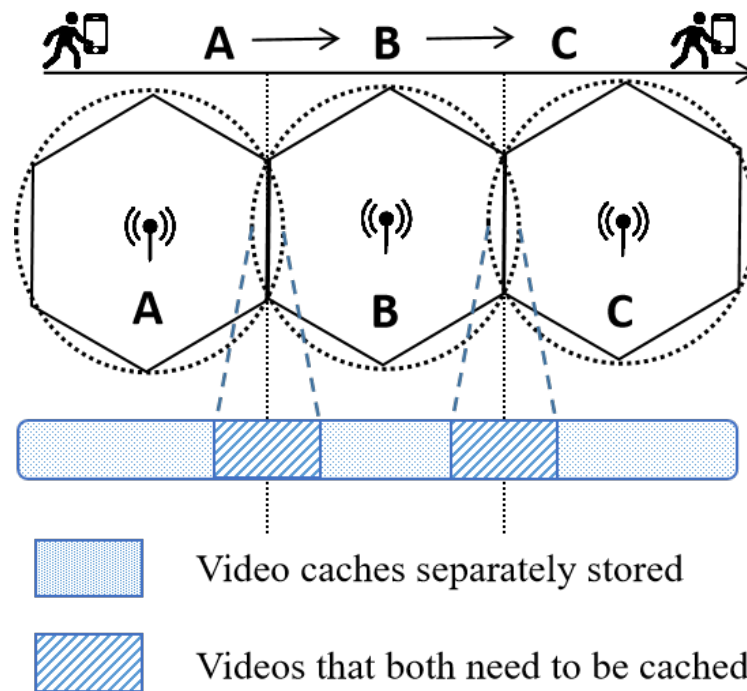
Figure 5. Video segmentation caching mechanism based on user's speed.

Next, we give the basic algorithm of the ECMSP strategy, which is mainly divided into two parts: grouping and optimization. Our algorithm is shown in Algorithm 1. We group MBSs and SBSs and cooperate to cache popular videos. First, set the buffer capacity $\mathcal{R}$ and $\mathcal{S}$ of different edge base stations, the total video popularity order p, and other attributes F_f of each video content. We initialize the communication radius d_M and d_N of MBSs and SBSs, the Poisson distribution intensity λ, the user's position coordinates and speed U_u, the background thermal noise power δ^2, etc. After a series of cycles and updating the caching strategy matrix, we obtain the optimal caching strategy matrix $\hat{c}_{nip}$ and the optimal total revenue function value $p_{total,max}$.

Algorithm 1 Edge caching strategy based on mobile speed and popularity

---

1: Getting the initialization matrix dimension
2: $beacon = 1$
3: **while** $beacon == 1$ **do**
4:     **while** $m < M$ **do**
5:         $m++$
6:         **while** $n < N$ **do**
7:             $n++$
8:             **while** $g < G_n(n)_{max}$ **do**
9:                 $g++$
10:                 Derive the cache probability $\hat{c}_{nip}$ with the manner in (9)
11:                 **if** $c_{nip} == 0$ **then**
12:                     Restricting the optimal caching strategy matrix
13:                     $c_{nip} = 0$
14:                 **end if**
15:                 Refresh matrix C and receive a new C_g
16:             **end while**
17:         **end while**
18:     **end while**
19:     **if** $C_g == C$ **then**
20:         $beacon = 0$
21:     **end if**
22: **end while**
23: Calculating the maximum value of total revenue function $p_{total,max}$ in (8)

---

4. Simulations and Results

In this section, we will evaluate the cache performance of our edge caching strategy in a mobile simulation network environment from different performance indicators.

4.1. Simulation Settings

4.1.1. Parameters' Settings

Based on the above caching strategy, we implement with our simulation through Python. The application scenario is defined as mobile video pre-caching in an edge computing network environment. In this simulated urban traffic environment, mobile users have different moving speeds due to different vehicles. These users browse a variety of videos through handheld mobile devices. By tracking the user's speed, we cached the video content in advance in the base station on its path, thereby reducing the time delay for users to obtain the video and the video playback jam, and improving the user experience quality. Only when the size of the video is large is there a greater possibility that the video playback will freeze. Due to the different popularity and bit rate of video content, we set the size of the video content package in the simulation experiment to 500 MB to 1000 MB. The mobile video behavior data set used in the experiment is selected from a part of real data collected by a video provider company [31]. We observe that the popularity of mobile video content follows a power-law distribution, as shown in Figure 6.
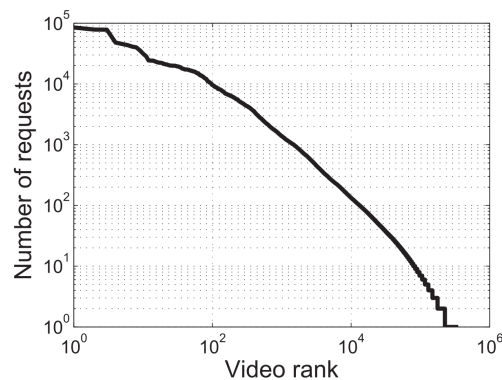
Figure 6. Mobile video popularity distribution.

According to the 4G-LTE communication standard [32], we set the wireless transmission rate of the uplink/downlink between the SBSs and the mobile user's handheld device to 50 Mbps, and the wired transmission rate among SBSs, MBSs and the central cloud server is 100 Mbps [33]. In a natural environment, we found that the coverage of the base station signal is affected by the density of buildings and population. In the simulation experiment, we set the service coverage radius of SBSs and MBSs to 150 m and 500 m, respectively. Furthermore, the experiment set up 7 MBSs and 72 SBSs in a cellular distribution mode. The users of handheld mobile devices have different speeds and follow the Poisson distribution. The detailed simulation parameter settings are shown in Table 1.

Table 1. Simulation parameters.

Parameters	Value
Simulation time (s)	200
Area (km $*$ km)	$2.5 * 2.5$
The number of MBSs	7
The number of SBSs	72
The number of mobile users	100–500
Cache size of central cloud (TB)	10
Cache size of MBS (TB)	1
Cache size of SBS (GB)	10
Coverage of MBS (m)	500
Coverage of SBS (m)	150
Mobile user's speed (m/s)	1–20
Video size (MB)	500–1000
r_{SU} (Mbps)	50
r_{MS} (Mbps)	100
r_{CM} (Mbps)	100
r_{CU} (Mbps)	100

4.1.2. Comparison Schemes

To more intuitively evaluate the caching performance of our proposed strategy (ECMSP), we compared it with the other three solutions:

- RPCC (Random Proactive Content Caching scheme), it sorts the content according to its popularity, and randomly caches highly popular content.
- MLCP (Machine Learning-based Content Prefetching scheme [26]), it uses machine learning algorithms to detect network traffic, and uses a multi-location and multi-factor prefetching scheme to pre-cache content in the VANET.
- CCMP (Cooperative Caching based on Mobility Predicting [27]), it selects the base station with the longest stay on the user path as the cache node.

4.1.3. Performance Metrics

- Transmission delay

  For video streaming services in a mobile edge computing environment, transmission delay is the most intuitive performance indicator that affects the quality of user experience, and it is also an essential requirement that reflects the performance of the caching strategy. In the experiment, the time delay here is expressed as the time interval from when the mobile user requests the video content to the time when the content is obtained, which mainly includes the process of request transmission, task processing, content transmission, etc.

- Cache Hit Ratio

  Since we are testing the performance of the edge caching scheme, we selected the cache hit ratio of SBSs as an indicator to reflect the performance of the cache, which can reflect the effectiveness of the video content pre-caching process.

4.2. Results Analysis

4.2.1. Average Delay

As shown in Figure 7a, we tracked 300 mobile users, and it illustrates the results of average delay with simulation time. It can be seen from the figure that the time delays of these four schemes all show a trend of rising first and then falling from the first second. At the beginning of the simulation, the caches in MBS and SBS are empty. In all schemes, the base station needs to ask for the central cloud server to serve mobile users' requests according to their respective strategies. Therefore, the average delay of each scheme has little difference, and it starts to increase with the increase of mobile user requests. As the simulation continues, the caching strategies of different schemes all begin to work and reduce the time delay to varying degrees. Among them, the average delay of the ECMSP proposed by us is the lowest compared to the other three schemes. Because RPCC randomly pre-caches video content with high popularity and does not consider the spatial aggregation of video content, it has a sizable average delay and poor experimental results. In addition, MLCP considers the video traffic distribution of the cache network but ignores the changes in traffic caused by user movement, so the pre-caching effect is relatively poor. The pre-caching strategy of CCMP based on PPM ignores the influence of different speeds of mobile users. This is because the user's moving speed results in different stay times in different areas. Furthermore, the peaks of each scheme appear at different times, and the earlier delay peaks have stronger dynamic adaptability, which proves that our ECMSP has the best dynamic adaptability and the strongest service capabilities.
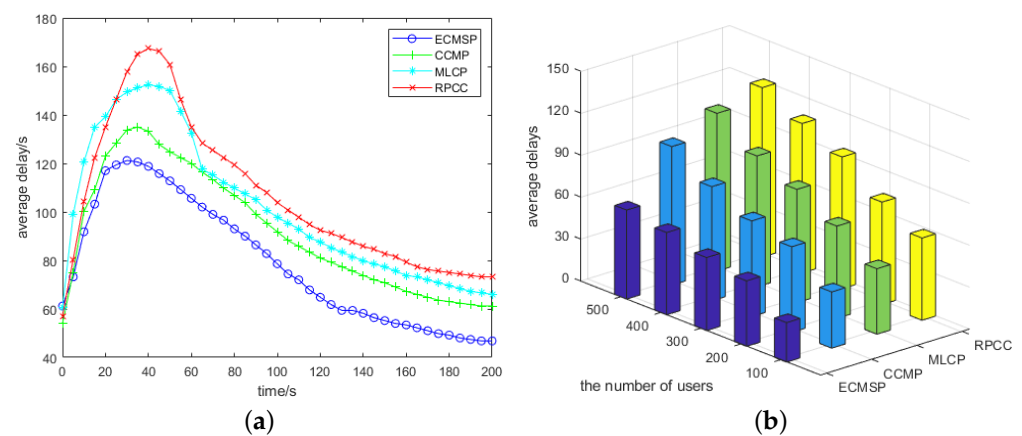

Figure 7. Comparison of time delay performance of four caching schemes: (**a**) average delay with the simulation time; (**b**) average delay with the number of the mobile users.

In Figure 7b, by increasing the number of mobile users, we found that the average delay of all caching schemes will increase, but our ECMSP still achieved the lowest latency

because, as mobile users increase, user requests for video traffic increase. On the one hand, base station connection handover and service interruption may increase. On the other hand, the increase in video traffic may exceed the capacity of the cache in the base station, and the cache is updated more frequently. The reason why the increase in mobile user requests has little impact on our strategy, and we believe this is mainly because of our comprehensive pre-cache decision-making mechanism and effective cache update strategy.

4.2.2. Cache Hit Ratio

Figure 8a shows the cache hit ratio results with 300 mobile users. The time when the inflection point of the curve in the figure appears illustrates the dynamic adaptability of different caching schemes. The inflection point of the ECMSP curve appears earliest, indicating that it first reaches the peak of the cache hit ratio. It can be seen from the figure that the cache hit ratio of RPCC is the lowest. This is because its cache is random and does not focus on hot videos. CCMP's cache hit ratio performance is better than MLCP. However, our ECMSP achieves the highest cache hit ratio. This is because ECMSP first caches highly popular video content in SBSs, and then caches video content of an appropriate size in the base station on its path according to the speed of the mobile users, thereby avoiding the problem of a decrease in cache hit ratio caused by the handoffs between base stations.
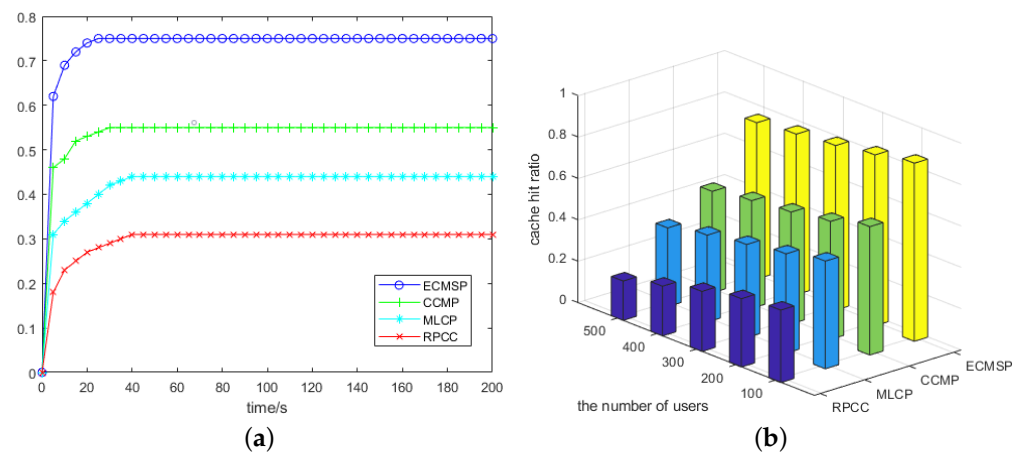

Figure 8. Comparison of cache hit ratio performance of four caching schemes: (**a**) cache hit ratio with the simulation time; (**b**) cache hit ratio with the number of mobile users.

As shown in Figure 8b, with the number of mobile users increasing, the cache hit ratio of RPCC, MLCP, and CCMP decreases, just as the analysis of Figure 7b, it is caused by the increase in requests for video content. However, the cache hit ratio of our ECMSP gradually stabilizes with the rise in the number of users, and has been maintained at a relatively high level, as it comprehensively considers the popularity of video content and the speed of mobile users, timely updates the cache, and can adapt to changes in the dynamic environment in time.

4.2.3. Change the Cache Capacity

Figure 9a,b illustrate the impact of changing the cache capacity of SBSs on the four schemes' average delay and cache hit ratio when the number of mobile users is 300. As we have seen, as the cache capacity of SBSs increases, the average latency of all schemes has been significantly reduced, especially the previous RPCC, which has a poor cache effect because it is less optimized than other schemes for caching strategies. For MLCP, CCMP, and ECMSP, although the average delay is also reduced to varying degrees, the effect is not very obvious. Nevertheless, our ECMSP still has the least average delay. Increasing the cache capacity of SBSs has a more significant impact on the cache hit ratio because the larger the cache capacity, the more video content can be cached. With the increase of SBSs cache capacity from 2 to 6 GB, our ECMSP cache hit ratio increased by 21%, and it has been

higher than the other three solutions. The experimental results can provide some valuable experience for improving the performance of the caching scheme.
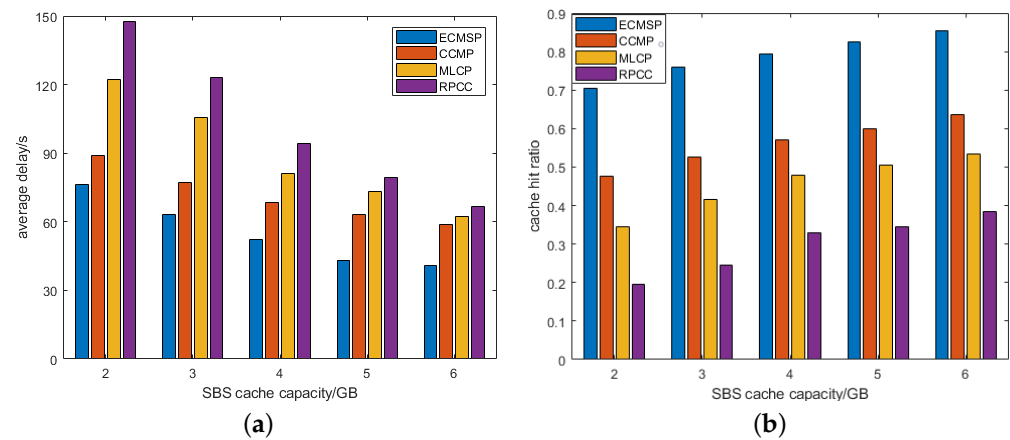
Figure 9. The impact of changing the cache capacity of SBSs on cache performance: (**a**) average delay; (**b**) cache hit ratio.

5. Discussion

This paper has proposed an edge caching strategy for mobile video streaming, combined with map navigation services, which can cache the corresponding content in advance on the base station on the user's route based on the user's moving speed and the popularity of the video content when the user's moving path is known. The purpose of this research work is to reduce the impact of the handoffs between base stations and service interruption during user movement, reduce the time delay for users to obtain video, improve the cache hit ratio of edge base stations, and improve the quality of user experience through caching strategies.

In order to evaluate the effectiveness of this strategy, we conducted a comparative experiment with the method in the latest literature. Although it was a simulation experiment, the data set of the video content in our experiment was selected from real data collected by a company engaged in the industry. By comparison, our proposed caching strategy has achieved better performance than other solutions in terms of average delay and cache hit ratio. The fly in the ointment is that only road models can be used in the simulation of mobile scenes, and the real factors of road traffic cannot be considered, such as the signal blocking by buildings or vehicles, and the influence between roadside signal sources. In addition, testing the performance of the caching strategy in the field will encounter other real challenges. On the one hand, the user's mobile information is obtained, which involves the user's security and privacy. A strategy is needed that can obtain the user's moving path in advance without infringing on the user's interests. On the other hand, some videos cannot be communicated between users due to copyright issues, which involves the intellectual property rights of the video content, and a specific strategy is needed to solve this problem. Therefore, the following other in-depth research needs to be carried out in actual cases to determine the effectiveness of the caching strategy.

In addition to the improvements in the experimental part, our future improvements will focus on the following. First of all, most of the current video content management and distribution technologies are based on traditional CDN systems, and there is no recognized solution for the research of mobile scenarios. Applying some mature technologies of CDN to mobile edge computing is one of our future research directions. Secondly, our caching strategy is based on the premise that the user's moving path is known; this requires the help of real-time map navigation services. The current user mobile location information is controlled by the operators related to map navigation, so a specific incentive mechanism is needed to obtain this information [34]. Furthermore, we consider optimizing the pre-caching design to refine the attributes of users, services and content, such as expanding

the pre-caching capacity in densely populated areas, or setting additional capacity for emergency tasks to ensure timely response; as some video content has prominent regional characteristics, we need to add regional attributes to the video to increase their cache priority in certain regions. In addition, it is also a good idea to apply machine learning algorithms to the cache update mechanism.

References

1. Update, I. *Ericsson Mobility Report*; Ericsson: Stockholm, Sweden, 2018.
2. Armbrust, M.; Fox, A.; Griffith, R.; Joseph, A.D.; Katz, R.H.; Konwinski, A.; Lee, G.; Patterson, D.A.; Rabkin, A.; Stoica, I.; et al. *Above the Clouds: A Berkeley View of Cloud Computing*; Technical Report, Technical Report UCB/EECS-2009-28; EECS Department, University of California: Berkeley, CA, USA, 2009.
3. Cisco, U. Cisco Annual Internet Report (2018–2023) White Paper. 2020. Available online: https://www.cisco.com/c/en/us/solutions/collateral/executive-perspectives/annual-internet-report/whitepaper-c11-741490.html (accessed on 26 March 2021).
4. Abbas, N.; Zhang, Y.; Taherkordi, A.; Skeie, T. Mobile edge computing: A survey. *IEEE Internet Things J.* **2017**, *5*, 450–465. [CrossRef]
5. Dai, X.; Wang, X.; Liu, N. Optimal scheduling of data-intensive applications in cloud-based video distribution services. *IEEE Trans. Circuits Syst. Video Technol.* **2016**, *27*, 73–83. [CrossRef]
6. Zahed, M.I.A.; Ahmad, I.; Habibi, D.; Phung, Q.V. Content caching in industrial IoT: Security and energy considerations. *IEEE Internet Things J.* **2019**, *7*, 491–504. [CrossRef]
7. Wang, R.; Kan, Z.; Cui, Y.; Wu, D.; Zhen, Y. Cooperative Caching Strategy With Content Request Prediction in Internet of Vehicles. *IEEE Internet Things J.* **2021**, *8*, 8964–8975. [CrossRef]
8. Gu, J.; Ji, Y.; Duan, W.; Zhang, G. Node Value and Content Popularity-Based Caching Strategy for Massive VANETs. *Wirel. Commun. Mob. Comput.* **2021**, *2021*, 5585896. [CrossRef]
9. Chen, Y.; Chen, S.; Chen, X. Efficient caching strategy in wireless networks with mobile edge computing. *Peer-to-Peer Netw. Appl.* **2020**, *13*, 1756–1766. [CrossRef]
10. Chunlin, L.; Zhang, J. Dynamic cooperative caching strategy for delay-sensitive applications in edge computing environment. *J. Supercomput.* **2020**, *76*, 7594–7618. [CrossRef]
11. Yang, R.; Guo, S. A Mobile Edge Caching Strategy for Video Grouping in Vehicular Networks. In Proceedings of the 2021 13th International Conference on Advanced Computational Intelligence (ICACI), Chongqing, China, 14–16 May 2021; pp. 40–45.
12. Mukerjee, M.K.; Hong, J.; Jiang, J.; Naylor, D.; Han, D.; Seshan, S.; Zhang, H. Enabling near real-time central control for live video delivery in CDNs. In Proceedings of the 2014 ACM conference on SIGCOMM, Chicago, IL, USA, 17–22 August 2014; pp. 343–344.
13. Xu, C.; Jia, S.; Wang, M.; Zhong, L.; Zhang, H.; Muntean, G.M. Performance-aware mobile community-based VoD streaming over vehicular ad hoc networks. *IEEE Trans. Veh. Technol.* **2014**, *64*, 1201–1217. [CrossRef]
14. Brodersen, A.; Scellato, S.; Wattenhofer, M. Youtube around the world: geographic popularity of videos. In Proceedings of the 21st international conference on World Wide Web, Lyon, France, 16–20 April 2012; pp. 241–250.
15. Xu, J.; Van Der Schaar, M.; Liu, J.; Li, H. Forecasting popularity of videos using social media. *IEEE J. Sel. Top. Signal Process.* **2014**, *9*, 330–343. [CrossRef]
16. Das, A.K.; Pathak, P.H.; Chuah, C.N.; Mohapatra, P. Contextual localization through network traffic analysis. In Proceedings of the IEEE INFOCOM 2014-IEEE Conference on Computer Communications, Toronto, ON, Canada, 27 April–2 May 2014; pp. 925–933.
17. Gitzenis, S.; Paschos, G.S.; Tassiulas, L. Asymptotic laws for joint content replication and delivery in wireless networks. *IEEE Trans. Inf. Theory* **2012**, *59*, 2760–2776. [CrossRef]
18. Wang, X.; Chen, M.; Taleb, T.; Ksentini, A.; Leung, V.C. Cache in the air: Exploiting content caching and delivery techniques for 5G systems. *IEEE Commun. Mag.* **2014**, *52*, 131–139. [CrossRef]
19. Passarella, A. A survey on content-centric technologies for the current Internet: CDN and P2P solutions. *Comput. Commun.* **2012**, *35*, 1–32. [CrossRef]

20. Ghabashneh, E.; Rao, S. Exploring the interplay between CDN caching and video streaming performance. In Proceedings of the IEEE INFOCOM 2020-IEEE Conference on Computer Communications, Toronto, ON, Canada, 6–9 July 2020; pp. 516–525.

21. Shabrina, W.E.; Sudiharto, D.W.; Ariyanto, E.; Al Makky, M. The Usage of CDN for Live Video Streaming to Improve QoS. Case Study: 1231 Provider. *J. Commun.* **2020**, *15*, 359–366. [CrossRef]

22. Jiang, W.; Feng, G.; Qin, S.; Liu, Y. Multi-agent reinforcement learning based cooperative content caching for mobile edge networks. *IEEE Access* **2019**, *7*, 61856–61867. [CrossRef]

23. Poularakis, K.; Iosifidis, G.; Argyriou, A.; Koutsopoulos, I.; Tassiulas, L. Caching and operator cooperation policies for layered video content delivery. In Proceedings of the IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications, San Francisco, CA, USA, 10–14 April 2016; pp. 1–9.

24. Guan, Y.; Zhang, X.; Guo, Z. Prefcache: Edge cache admission with user preference learning for video content distribution. *IEEE Trans. Circuits Syst. Video Technol.* **2020**, *31*, 1618–1631. [CrossRef]

25. Su, Z.; Hui, Y.; Xu, Q.; Yang, T.; Liu, J.; Jia, Y. An edge caching scheme to distribute content in vehicular networks. *IEEE Trans. Veh. Technol.* **2018**, *67*, 5346–5356. [CrossRef]

26. Luo, G.; Yuan, Q.; Zhou, H.; Cheng, N.; Liu, Z.; Yang, F.; Shen, X.S. Cooperative vehicular content distribution in edge computing assisted 5G-VANET. *China Commun.* **2018**, *15*, 1–17. [CrossRef]

27. Yao, L.; Chen, A.; Deng, J.; Wang, J.; Wu, G. A cooperative caching scheme based on mobility prediction in vehicular content centric networks. *IEEE Trans. Veh. Technol.* **2017**, *67*, 5435–5444. [CrossRef]

28. Yang, X.; Fei, Z.; Zheng, J.; Zhang, N.; Anpalagan, A. Joint multi-user computation offloading and data caching for hybrid mobile cloud/edge computing. *IEEE Trans. Veh. Technol.* **2019**, *68*, 11018–11030. [CrossRef]

29. Breslau, L.; Cao, P.; Fan, L.; Phillips, G.; Shenker, S. Web caching and Zipf-like distributions: Evidence and implications. In Proceedings of the IEEE INFOCOM'99. Conference on Computer Communications. Proceedings. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies. The Future is Now (Cat. No. 99CH36320), New York, NY, USA, 21–25 March 1999; Volume 1, pp. 126–134.

30. Li, C.; Lin, L.; Zuo, W.; Wang, W.; Tang, J. An approach to streaming video segmentation with sub-optimal low-rank decomposition. *IEEE Trans. Image Process.* **2016**, *25*, 1947–1960. [CrossRef]

31. Ma, G.; Wang, Z.; Zhang, M.; Ye, J.; Chen, M.; Zhu, W. Understanding performance of edge content caching for mobile video streaming. *IEEE J. Sel. Areas Commun.* **2017**, *35*, 1076–1089. [CrossRef]

32. Xu, Z.; Li, X.; Zhao, X.; Zhang, M.H.; Wang, Z. DSRC versus 4G-LTE for connected vehicle applications: A study on field experiments of vehicular communication performance. *J. Adv. Transp.* **2017**, *435*. [CrossRef]

33. Molina-Masegosa, R.; Gozalvez, J. LTE-V for sidelink 5G V2X vehicular communications: A new 5G technology for short-range vehicle-to-everything communications. *IEEE Veh. Technol. Mag.* **2017**, *12*, 30–39. [CrossRef]

34. Wu, R.; Tang, G.; Chen, T.; Guo, D.; Luo, L.; Kang, W. A Profit-aware Coalition Game for Cooperative Content Caching at the Network Edge. *IEEE Internet Things J.* **2021**, 1. [CrossRef]