Multimedia Tools and Applications (2020) 79:25535–25549
https://doi.org/10.1007/s11042-020-09208-7

A novel video delivery mechanism for caching-enabled networks

Zhimin Qi[1]

Received: 8 July 2019 / Revised: 20 April 2020 / Accepted: 11 June 2020 /
Published online: 4 July 2020

Abstract
The caching-enabled networks, especially Information-Centric Networking (ICN), have attracted much attention from some global research communities, and the corresponding achievements have been highlighted in the field of video delivery. This paper studies more effective video delivery mechanism in order to guarantee better network performance based on two conceptions, i.e., the intermediate caching of popular video segments and the light delivery of identity object. At first, we evaluate the popularity under the dynamic environment, including descending modelling, ascending modelling and period modelling. Then, we propose identity-based light video delivery mechanism; in particular, we introduce Data Plane Development Kit (DPDK) to ensure that all video segments have the same size and to accelerate video transmission by bypassing the kernel. Finally, the simulation experiments are made based on the real YouTube dataset over CERNET network topology, and the results demonstrate that the proposed video delivery mechanism outperforms two the-state-of-the-art mechanisms in terms of cache hit ratio, routing hop count, delivery delay and network load.

Keywords Caching-enabled networks · Video delivery · Popularity modelling · Light mechanism · DPDK

1 Introduction

Recently, the characteristic of in-network caching has been accepted and canonized by some prevalent networking paradigms, such as Peer-to-Peer Networking (P2PN) [1], Content Delivery Networks (CDN) [2], Information-Centric Networking (ICN) [3], Hybrid ICN (HICN) [4] and IPv6 Content Networking (6CN) [5]. Among them, (i) P2PN, CDN and ICN are proposed by the academic communities, while HICN and 6CN are proposed by the industrial community, i.e., Cisco; (ii) P2PN, CDN and 6CN are overlay solutions, while

✉ Zhimin Qi
qizhimin1983@sina.com

[1] School of Media, Jilin Normal University, Siping, 136000, China

ICN and HICN are the built network-layer solutions; (iii) CDN has the most mature and extensive market application (e.g., Akamai [6]), while ICN has the strongest academic value (e.g., ACM international conference on ICN [7]). Under such context, more and more content providers have been emerging, such as YouTube, NetFlix, iQIYI and Tencent, and they accelerate the delivery of videos to edge users by leveraging in-network caching. In spite of this, the current video delivery schemes have some obvious limitations, such as redundant delivery, high network load and long delivery delay. Therefore, the more effective video delivery mechanism for caching-enabled networks should be investigated in order to guarantee better network performance.

According to the report in [8], we know that 83% watching time concentrates on 1% videos, which indicates that the popular (or hot) videos only account for a very small percentage. Furthermore, [9] shows that, a video is divided into a number of parts with same duration whist only two or three parts are watched with ultra-high frequency, which suggests that only a few segments of a popular video rather than the whole video are popular. Given this, Origin Server (OS) only need deliver the hot segments towards its downstream at the same time they are cached at the intermediate Content Routers (CRs), which can respond to users quickly. Another point on the subsequent video delivery, if the downstream has the cached hot segments, it is not necessary for OS and even CRs to downwardly deliver the video entity; instead, the corresponding ID can be regarded as the delivery object in order to reduce the redundant traffic transmission. Therefore, when CR judges its downstream has the hot satisfaction for the current request, the core idea of video delivery should be that the designated video name (packet header) irrespective of the requested data (payload) is delivered.

Regarding the core idea, it is summarized as two points, i.e., the intermediate caching of popular video segments and the light delivery of ID object. On one hand, although there are many studies on the former, they usually adopt Least Recently Used (LRU) as the replacement strategy to hold the popular video segments but do not thoroughly pay attention to the evaluation of popularity which is considerably important because the popularity is dynamic change. On the other hand, to our best of knowledge, the light delivery of ID object is the first proposal in this paper, which has the significant optimization effect on video delivery. Based on the above statements, the major contributions of the Novel Video Delivery mechanism for Caching-enabled networks (NVDC) in this paper are concluded as follows.

- The popularity is evaluated based on the dynamic environment, which includes three parts, i.e., descending modelling, ascending modelling and period modelling.
- The Data Plane Development Kit (DPDK) is used to ensure that all videos segments have the same size and further to accelerate video transmission by bypassing the kernel.
- The light video delivery scheme is proposed, in which if the downstream has the popular video segment, its corresponding ID is delivered; otherwise, the entity is delivered.
- The simulation is driven by the real YouTube dataset, and the experiments evaluate cache hit ratio, routing hop count, delivery delay and network load.

The remaining of this paper is structured as follows. Section 2 reviews and compares the related work. Section 3 presents the system framework of NVDC. In Section 4, the video caching scheme is determined. Section 5 devises the video delivery mechanism including the acceleration of video transmission and the details of delivery. Section 6 evaluates the proposed NVDC and finally Section 7 concludes this paper.

2 Related work

In recent years, a number of content delivery mechanisms have been investigated. As a matter of fact, some of them can also be suitable for the special video delivery scenario. Thus, this section reviews including but not limited to video delivery from four aspects, i.e., delivery at P2PN, delivery at CDN, delivery at ICN, and delivery at 6CN/HICN.

There are some content delivery mechanisms designed for P2PN. For example, in [10], an efficient content delivery scheme that used network coding was proposed to share large files, in which peers requested file blocks from multiple server nodes and servers sent blocks to multiple receivers, providing the multipoint-to-multipoint communication. In [11], a secure content delivery strategy was proposed to solve the problem of key leakage, which ensured that only the legitimate users could access the content. In [12], a P2PN assisted model for streaming Video on Demand (VoD) contents was presented, which took advantages of the unused uplink and storage capacity to serve requests of other clients. In [13], the caching capacities of end-users was exploited to optimize content delivery. In [14], a fully distributed in-network caching protocol (with chunk availability, popularity and peer distance consideration) for P2P-like content chunks was proposed to reduce P2P based traffic load and improve delivery performance. In [15], a smart recommender for enhancing content delivery by taking into account the analysis of user's behavior was devised. In [16], by utilizing the principal-agent model of incentive theory, a delivery scheme was studied to support the operations of online file exchange service.

For CDN, there are also some representative delivery mechanisms. For example, in [17], a recursive hierarchical push-based cooperative replica replacement mechanism was proposed to improve delivery efficiency and it was formalized based on the economic mechanism design theory. In [18], a delivery strategy was designed based on several net constraints such as system load, network infrastructure, user satisfaction and migration cost, where a cache placement protocol was explored to support the cache migration. In [19], a network was divided into several domains, and each domain had a cluster address. Based on this, when some contents had the same destination, they were aggregated and only a content was delivered by monitoring the paths. In [20], a novel approach where an Internet Service Provider (ISP) in an ISP-CDN collaboration actively redirected end users to the selected server was proposed. It surrogated servers via a combination of Software-Defined Networking (SDN) and Network Functions Virtualization (NFV) techniques. In [21], a general AI-defined attention network was proposed to make predictions (a crucial prerequisite for subsequent policies) in order to provide the optimal resource scheduling for CDN. In [22], the competition between ISPs where users required heterogeneous contents was analyzed to adjust content delivery by considering net neutrality and paid prioritization.

There exist many research achievements related to content delivery in ICN. For example, in [23], a seamless content delivery scheme was proposed to solve node mobility problem by using flow mapping agent which managed information connection between CRs. In [24], the inventory model of supply chain management in logistics was used to formulate the process of content delivery. In [25], a relay-based multipoint content delivery in wireless context was proposed by taking advantage of inherent content awareness, where a multi-objective optimization framework and a low-complexity heuristic algorithm were provided for relay selection and transmission rate assignment. In [26], an SDN-based content delivery scheme was devised, in which SDN took charge of the guiding of content delivery. In [27], the analytical model was developed to characterize the multi-path content delivery with the

parallelized chunk requests. In [28], inspired by the natural ant behaviors, the distributed and parallel content delivery scheme was devised, in which the delivery process was completed in a self-organized way. In [29], a multicast architecture which integrated the stateless forwarding to partition the multicast tree and establish cost-efficient branches was proposed to effectively deliver content to the multiple end hosts. In [30], the network coding technique was adopted to improve in-network cache utilization so that network throughput and efficiency were increased, avoiding the duplicate and unproductive chunk delivery while transferring disjoint segments along the multiple paths. In [31], a testbed was built for the special media content delivery, which was implemented over Long Term Evolution (LTE) ratio access networks.

To the best of our knowledge, 6CN and HICN are two latest networking paradigms and there have not yet some related researches on content delivery. Although the above reviewed publications have optimization effect on content delivery to some extent, they always have some limitations, such as redundant delivery, high network load and long delivery delay. Different from them, in this paper, we give a novel video delivery mechanism, which can reduce the transmitted traffic with the maximum level. In addition, from the perspective of engineering implementation, we use DPDK to accelerate video deliery, which is not involved by the other related delivery mechanisms. Particularly, we do simulation based on the real YouTube dataset which contains the sufficient short videos, which makes the experiments more convincing.

3 System framework

As depicted in Fig. 1, the proposed NVDC consists of three major modules, i.e., Popularity Evaluation Module (PEM), DPDK Usage Module (DUM) and Light Delivery Module (LDM). Among them, PEM is used to evaluate the popularity of video segment and cache the relatively popular video segments at CR. LDM is used to deliver the ID of segment if the downstream has the cached segment, whilst to deliver the segment otherwise. DUM is used to ensure that all segments have the same size and to accelerate video transmission

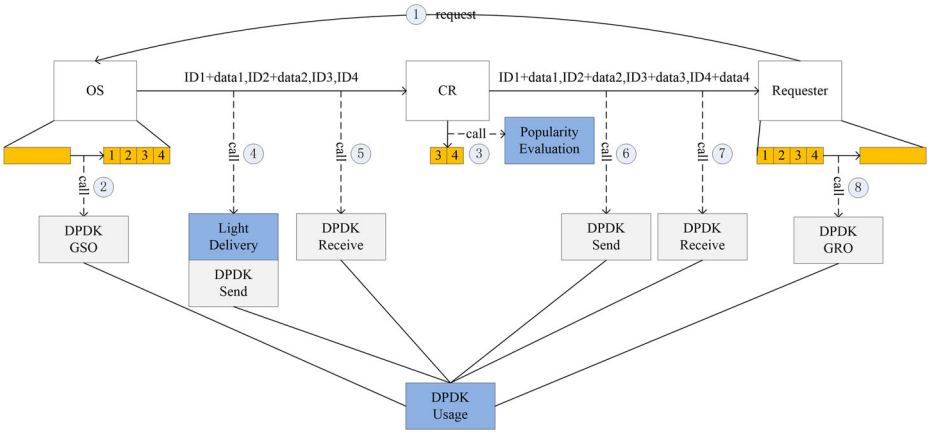

Fig. 1 The system framework of NVDC

by bypassing the kernel. In particular, DUM is composed of four sub-modules, i.e., DPDK Generic Segmentation Offload (GSO) that slices the whole video into a number of segments, DPDK send that bursts some segments to the network at once, DPDK receive that bursts some segments from the network at once, and DPDK Generic Receive Offload (GRO) that combines a number of segments into a whole video. For OS, the first and the second sub-modules are performed; for the intermediate CR, the second and the third sub-modules are performed; for the requester, the third and the last sub-modules are performed.

In addition, this paper considers one video segment as a packet that consists of two parts, i.e., ID and data. Among them, the ID is similar to content name in ICN and similar to the sequence number in TCP/IP, and it can be regarded the header of packet; the data can be regarded as the payload of packet. For the three roles (i.e., OS, CR and requester) and three modules (i.e., PEM, DUM and LDM) in Fig. 1, the simple workflow is described as follows.

- Step1. The requester sends a video request to OS.
- Step2. OS provides a matched video and calls DPDK GSO to slice the video into four packets, numbered as 1, 2, 3 and 4 respectively.
- Step3. The other requesters maybe have requested the video, where packet-3 and packet-4 are the popular segments, which are cached at CR, by calling PEM.
- Step4. OS knows that its downstream has packet-3 and packet-4, thus it calls LDM to deliver ID1+data1, ID2+data2, ID3 and ID4, where DPDK send is also called to improve delivery efficiency.
- Step5. CR calls DPDK receive to handle the delivered ID and data.
- Step6. CR knows that its downstream has no the corresponding packets, thus it calls DPDK send to directly deliver ID1+data1, ID2+data2, ID3+data3 and ID4+data4.
- Step7. The requester calls DPDK receive to handle the delivered ID and data.
- Step8. The requester calls DPDK GRO to combine these packets into a whole video, that is, the ID is deleted and the corresponding data is reserved.

4 Video caching

Because CR is subjected to the limited cache capacity, it has to cache the popular video segments rather than accommodating all videos. Therefore, the evaluation of popularity is considerably important. Even so, the traditional method for this usually adopts LRU that evaluates the popularity based on access frequency, which is the discrete evaluation method and cannot be suitable to the dynamic environment. In this paper, we use the continuous method to evaluate the popularity, including descending modelling, ascending modelling and period modelling.

4.1 Descending modelling

Suppose that $p(t)$ is the popularity at time t and that $\rho(t)$ is the attenuation factor of popularity, and we have three Constraint Conditions (CCs) on the descending modelling as follows. CC1: when a video or a segment does not be accessed for the transient time, the corresponding popularity becomes descending over time, that is, $p(t)$ is a decreasing function. CC2: when a video or a segment does not be accessed for along time, the corresponding popularity is 0. CC3: as time goes on, the popularity attenuates faster and faster, that is, $\rho(t)$ is a increasing function.

Inspired by the alcohol volatilization, we have

$$\begin{cases} \frac{\partial p(t)}{\partial t} = -\theta p(t) \\ p(t)_{|t=0} = p_0 \end{cases} \tag{1}$$

where p_0 is the initial popularity of $p(t)$ and θ is a positive constant. By solving the differential (1), we have

$$p(t) = p_0 e^{-\theta t} \tag{2}$$

$$\rho(t) = -\theta p_0 e^{-\theta t} < 0 \tag{3}$$

$$\lim_{t \to +\infty} p(t) = \lim_{t \to +\infty} p_0 e^{-\theta t} = 0 \tag{4}$$

According to (3), we know that $p(t)$ is a descending function, which indicates that CC1 is satisfied. According to (4), we know that $p(t)$ approaches 0 when t is the infinite, which indicates that CC2 is satisfied. Furthermore, we have

$$\frac{\partial \rho(t)}{\partial t} = -\theta^2 p_0 e^{-\theta t} > 0 \tag{5}$$

which means that $\rho(t)$ is a strictly increasing function. In other words, CC3 is satisfied.

4.2 Ascending modelling

If a video or a segment is accessed frequently, the corresponding popularity becomes ascending. Inspired by the object endothermy model, we have

$$Q = cm\Delta t \tag{6}$$

where Q is the heat that the object absorbs, c is the specific heat, m is the object's mass, and Δt is the temperature difference. Mapping them to the popularity, and we have that, Q is the popularity increment, m is the size of segment, Δt is the time difference, and c has different settings according to the application type.

Putting (6) into (2), and we have the comprehensive popularity evaluation function as follows.

$$p(t) = p_0 e^{-\theta t} + cm\Delta t \tag{7}$$

4.3 Period modelling

According to the analysis in [32], it is impossible that the popularity keeps ascending; instead, the popularity reaches a maximal value and only keeps a short period, which means that the popularity has the time-validity and temporality. For example, in terms of a recently released video, at the initial several weeks, its popularity keeps ascending; at the follow-up several weeks, its popularity keeps descending; after a number of weeks, its popularity approaches 0. Therefore, it requires to analyze the period of popularity, i.e., the modelling of t in (7), so that the caching time of a video or a segment at CR can be determined well.

Consider that the ascending rate of popularity cannot be larger than the given γ, and we have the boundary equation as follows.

$$\frac{p(t) - p(t-1)}{p(t-1)} = \gamma \tag{8}$$

Putting (7) into (8), and we have

$$p_0 e^{-\theta t} + cmt = \gamma + 1 \tag{9}$$

which means that t cannot be expressed by an analytical solution. Thus, regarding the determination of t, we only give its numerical solution when the simulation is made.

In summary, this section builds three models. Among them, the first one means that the popularity becomes descending over time; the second one means that the popularity becomes ascending due to the frequent access; and the last one means that the popularity has the temporality.

5 Video delivery

In this section, we present the light video delivery mechanism, i.e., NVDC. However, to improve the cache utilization rate at the most extent, it is indispensable to ensure that all cached segments have the same size. To this end, DPDK [33] is introduced for this matter. Furthermore, DPDK also has the function of accelerating video transmission by bypassing the kernel.

5.1 DPDK usage

As described in Section 3, the usage of DPDK involves DPDK GSO, DPDK send, DPDK receive and DPDK GRO. Among them, DPDK send and DPDK receive are very simple due to the corresponding operations can be completed by the built-in functions. To be specific, DPDK send is implemented by calling the built-in function rte_tx_burst, while DPDK receive is implemented by calling the built-in function rte_rx_burst. In addition, the combination based on GRO is only an engineering implementation and there are also many application cases for it. Thus, we in this paper do not give its discussion. Although the segmentation based on GSO is also an engineering implementation, it is only suitable for the standard header, i.e., TCP header, IP header and Ethernet header. As a matter of fact, the ID is not the standard header and cannot be resolved by the network. Therefore, the implementation of ID should be put behind the standard header, and the segmentation process includes the following four main operations. (i) A standard header is added for the delivering video; (ii) the built-in function rte_gso_segment is called to do segmentation, in which each segment consists of the standard header and payload; (iii) ID is inserted behind the standard header, and a new packet including the standard header, ID and payload is constructed in logical; and (iv) these new packets are put into a specified mbuf to prepare for sending.

The pseudo-code of GSO-based segmentation is described in **Algorithm 1**. Therein, videombuf is used to load the entire video; segmbuf is used to load the sliced standard packets; outmbuf is used to load the reconstructive packets with ID; and no_segs is the number of segments for each time. In particular, line 1 means the initialization of environment; line 2 means the creation of DPDK pool; lines 4-7 mean that it requires the multiple mbufs to

load the video; line 8 means the addition of the standard header for the video; line 9 means the operation of segmentation; and line 10-15 mean the reconstruction of new packet.

Algorithm 1 GSO-based segmentation.

```
Input: struct rte_mempool *mbuf_pool;
       struct rte_mbuf *videombuf[], *segmbuf, sheadermbuf, *IDmbuf;
       int packet_size, no_segs;
Output: struct rte_mbuf outmbuf;
01. rte_eal_init(argc,argv);
02. mbuf_pool=rte_pktmbuf_pool_create();
03. Pu the video into videombuf;
04. if the size of video is larger than 2048B, then
05.    Multiple videombufs are used;
06. else
07.    Only one videombuf is used;
08. rte_pktmbuf_chain(sheadermbuf,videombuf);
09. rte_gso_segment(videombuf,default&packet_size,segmbuf,no_segs);
10. for each segment, do
11.    *p=the header of segment;
12.    *q=the payload of segment;
13.    rte_pktmbuf_chain(p,IDmbuf);
14.    rte_pktmbuf_chain(IDmbuf,q);
15. endfor
16. outmbuf=segmbuf;
```

5.2 Light delivery

For the caching-enabled networks, each CR should be equipped with two tables, i.e., Delivery Information Table (DIT) and Content Store (CS), as shown in Fig. 2. DIT consists of three fields, i.e., timer and destination, which means that some video segment is delivered to some destination. Let T denote the timer duration, and the timer is set as T upon the delivery information entry is generated. CS consists of three fields, i.e., ID, data and timer, which means that some CR has some video segment and the survival time is T. In particular, OS is only equipped with DIT because it has owned all video segments.

With the help of DIT and CS, the core idea of light delivery mechanism is summarized as follows: if the downstream has some packet by checking DIT, the packet is stripped and it only remains the ID which is delivered; if CR receives ID but the downstream has no the corresponding packet, the data of the packet is loaded behind the ID of the packet by checking CS. Consider the delivering of packet-x (IDx+datax) from CR_i to CR_j, and the pseudo-code of light delivery is described in **Algorithm 2**. Therein, lines 2-12 involve the situation where CR receives packet-x while lines 13-22 involve the situation where CR only

Fig. 2 The structures of PIT and CS

receives IDx. Especially for line 20, it means that the received IDx has to be dropped when both CR_i and CR_j have no packet-x.

Algorithm 2 Light delivery.

Input: packet-x, IDx, datax, CR_i, CR_j;
Output: Null;
01. CR_i resolves the current packet;
02. **if** CR_i receives IDx+datax, **then**
03. **if** CS_i has packet-x, **then**
04. **if** DIT_i has IDx, **then**
05. Strip packet-x and deliver IDx to CR_j;
06. **else**
07. Deliver packet-x to CR_j;
08. **endif**
09. **else**
10. Cache packet-x;
11. Goto lines 4-8;
12. **endif**
13. **else**//CR_i receives IDx
14. **if** DIT_i has IDx, **then**
15. Deliver IDx to CR_j;
16. **else**
17. **if** CS_i has packet-x, **then**
18. Load datax behind IDx and deliver packet-x to CR_j;
19. **else**
20. Drop the received IDx;
21. **endif**
22. **endif**
23. **endif**

6 Performance evaluation

The simulation is driven by the real YouTube dataset, which is collected from a campus network measurement [32] for one day. The dataset contains 18751 requests and 13764 short videos. However, the dataset does not present the concrete network topology. Therefore, we select CERNET with 36 nodes and 43 edges [34] as the simulation topology, as shown in Fig. 3. By analyzing the distribution law of the dataset, we map 18751 requests and 13764 short videos to CERNET network topology. In addition, we simulate to set up CERNET network topology based on 36 personal hosts at the machine room from Tsinghua University, where each host is installed by DPDK-18.05 and allocated cache capacity by Docker.

Furthermore, we compare the proposed NVDC with two the-state-of-the-art mechanisms, i.e., AI based content delivery in CDN [21] and Network Coding based video delivery in ICN [30], AICDN and NCICN for short respectively. Meanwhile, Average Cache Hit Ratio (ACHR), Average Routing Hop Count (ARHC), Average Delivery Delay (ADD) and Average Network Load (ANL) are considered as four evaluation metrics. In terms of the request testing, we design five groups of requests by using different time periods in chronological order, i.e., 200, 400, 600, 800 and 1000, where these requests are non-overlapping.

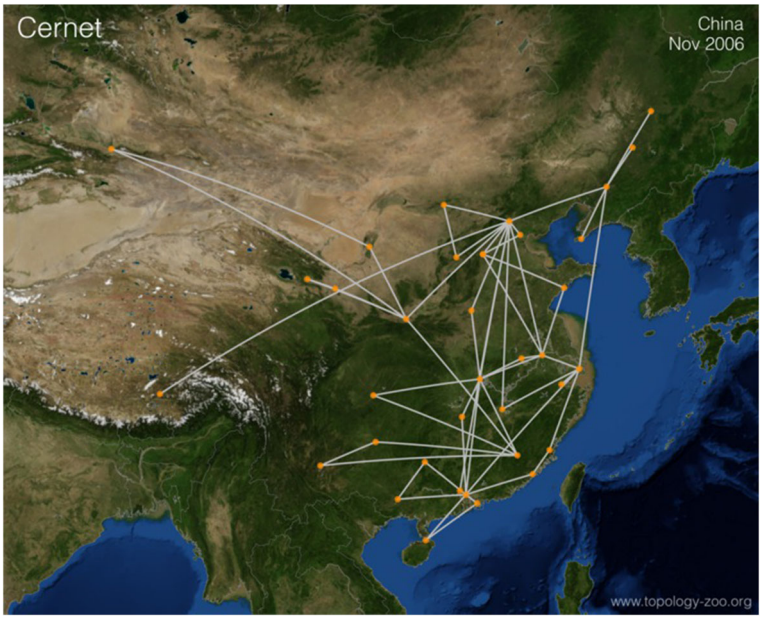

Fig. 3 CERNET network topology with 36 nodes and 43 edges used for simulation

Moreover, there are some simulation parameters including YouTube, DPDK, NVDC and CERNET. In terms of those involved in YouTube, which are obtained by the inherent dataset. In terms of those involved in CERNET, which are set according to the real network environment. For those involved in DPDK and NVDC, many groups of simulations under different settings are made and the most suitable combination is determined. To be specific, all parameters are set in Table 1.

Table 1 Simulation parameters

Parameter	Setting	Ownership
Collection Duration	24h	YouTube
The size of dataset	166GB	YouTube
The number of videos	13764	YouTube
The number of requests	18751	YouTube
The number of Hugepage	2048	DPDK
The size of Hugepage	2MB	DPDK
The packet_size	2500B	DPDK
T	4mins	NVDC
The size of cache	80MB	NVDC
The number of simulations	30	NVDC
The number of nodes	36	CERNET
The number of edges	43	CERNET
The network bandwidth	10Gb/s	CERNET

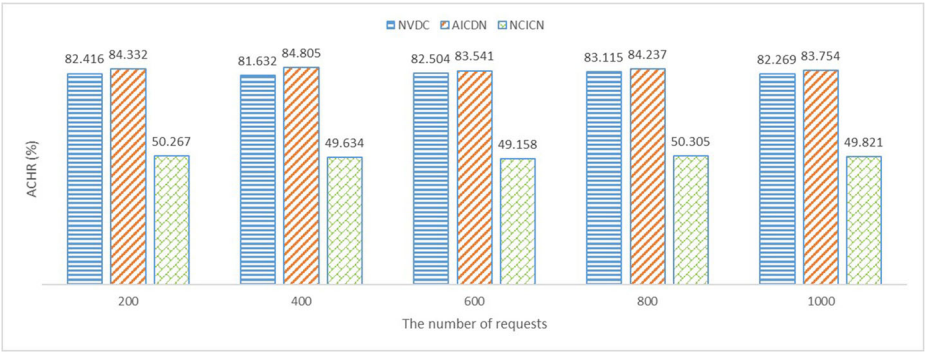

Fig. 4 ACHRs for three mechanisms

6.1 Cache hit ratio

The cache hit ratio is defined as $Num_{hit}/Num_{success}$, where Num_{hit} is the number of requests which are satisfied by CS and $Num_{success}$ is the successful number of requests. ACHRs for NVDC, AICDN and NCICN are reported in Fig. 4. We observe that AICDN has the highest ACHR, followed by NVDC and NCICN, about 84%, 82% and 50% respectively. Both NVDC and AICDN pay attention to the popularity and schedule the hot video segments to satisfy the users' requests. Although NCICN uses the technique of network coding and improves the cache utilization ratio, it does not effectively cache the hot video segments at the appropriate CRs. As a result, NCICN has the lowest ACHR. AICDN adopts the technique to predict the hot video segments and make the optimal resource scheduling, thus it can obtain the relatively hith ACHR. In spite of this, by reviewing the results, we can observe that ACHRs of NVDC and AICDN have litter difference. This is because NVDC has a special module, i.e., PEM to evaluate the popularity so that the hot video segments can be cached at the appropriate CRs for the corresponding time period. In other words, NVDC not only keeps a close watch on the hot video segments but also effectively plans the stayDuration, thus it can obtain the considerably high ACHR.

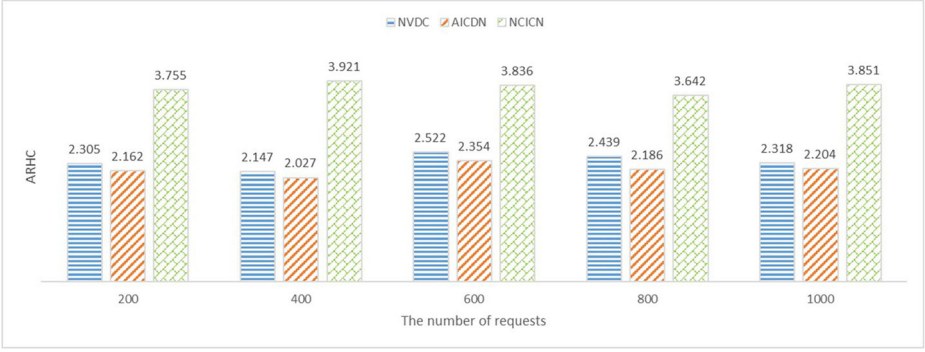

Fig. 5 ARHCs for three mechanisms

6.2 Routing hop count

The routing hop count is defined as the difference between the traversed number of CRs and one. ARHCs for NVDC, AICDN and NCICN are reported in Fig. 5. We observe that AICDN has the smallest ARHC, i.e., the best performance in terms of ARHC, followed by NVDC and NCICN. For testing many requests, if the subsequent requests can be satisfied by the intermediate CRs, the corresponding ARHC has a good expectation result. In other words, if a mechanism has high ACHR, its ARHC is small. As Section 6.1 reported, AICDN has the highest ACHR, thus it always has the smallest ARHC. Similar to the above analyzed reasons, ARHC of NVDC is close to that of AICDN.

6.3 Delivery delay

The delivery delay is defined as the difference between the timepoint when the requested video is delivered and that when it is obtained. ADDs for NVDC, AICDN and NCICN are reported in Fig. 6. We observe that NVDC has the smallest ADD, followed by AICDN and NCICN, about 50ms, 75ms and 95ms respectively, which further indicates that NVDC has the absolute advantage in terms of ADD. At first, AICDN adopts the technique of AI and NCICN adopts the technique of network coding, which consumes more computation time than NVDC because NVDC is only a heuristic mechanism. Secondly, NVDC is a light delivery mechanism, which only delivers ID rather than the video segment entity if the downstream has the cached video segment. In fact, such condition is very common, which can be conducted by ACHR. Thus, NVDC transmits much smaller traffic than AICDN and NCICN, and has smaller transmission time. At last, NVDC has a special module, i.e., DUM to accelerate video transmission while AICDN and NCICN does not involve this, thus for the same object, NVDC can transmit the object with smaller time than AICDN and NCICN. In summary, NVDC has smaller ADD than AICDN and NCICN. With respect to the two baselines, although the prediction based on AI in AICDN consumes more computation than the usage based on network coding in NCICN, AICDN has the absolute advantages than NCICN in terms of ACHR and ARHC. With such consideration, AICDN has smaller ADD than NCICN.

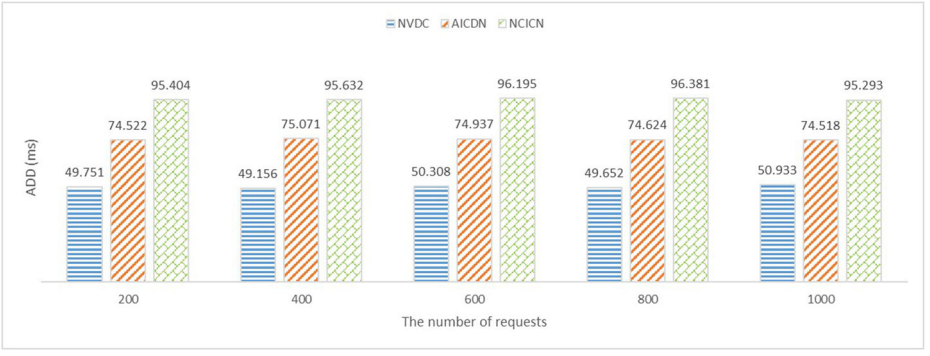

Fig. 6 ADDs for three mechanisms

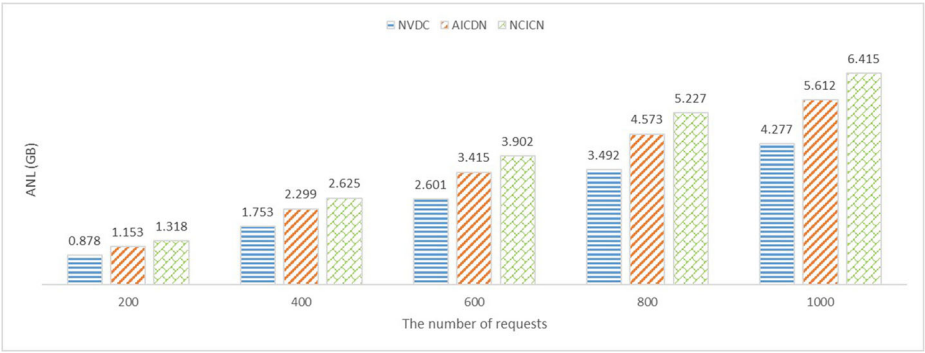

Fig. 7 ANLs for three mechanisms

6.4 Network load

The network load is defined as the total transmitted traffic volume when the corresponding number of requests are satisfied. For different groups of requests, we report the total traffic volumes when there is no any optimization mechanism in Table 2.

ANLs for NVDC, AICDN and NCICN are reported in Fig. 7. We observe that three mechanisms all have the optimization effect on network load, while the effect of NVDC is the best, followed by AICDN and NCICN. For example, when the number of requests is 200, we compute the corresponding saved traffic volume as about 52%, 37% and 28% respectively. According to Fig. 4, we know that NVDC has the considerably high ACHR, which implies that NVDC delivers ID in most situations due to the design of light delivery mechanism. As a result, NVDC has the smallest ANL. For AICDN and NCICN, the former has higher ACHR than the latter, which causes that the former transmits smaller traffic and thus has smaller ANL than the latter.

6.5 Discussion

By reviewing the experimental results reported in Sections 6.1–6.4, we know that NVDC has the best performance on ADD and ANL, and the second best performance on ACHR and ARHC. Although NVDC does not reach the best performance on ACHR and ARHC, it has the close performance to AICDN. In short, by evaluating the above four metrics, we think that the proposed NVDC has the best performance. In addition, for the designed three modules, i.e., PEM, DUM and LDM. According to the experimental results on ACHR, ADD and ANL respectively, it is obvious that PEM, DUM and LDM show good optimization effect.

Table 2 The total traffic volumes without any optimization mechanism

The number of requests	200	400	600	800	1000
Data volume	1.83GB	3.65GB	5.42GB	7.26GB	8.91GB

7 Conclusions

The characteristic of in-network caching has been accepted by some prevalent networking paradigms, which contributes to the caching-enabled networks, such as CDN, ICN and 6CN. The video delivery is one of the hottest research points and has attracted much attention from some global research communities. In spite of this, the current video delivery mechanisms have some obvious limitations, such as redundant delivery, high network load and long delivery delay, which need to be optimized. Therefore, this paper investigates more efficient video delivery mechanism to guarantee better network performance, which consists three modules, i.e., PEM, DUM and LDM. To be specific, PEM evaluates the popularity of video segment, including descending modelling, ascending modelling and period modelling. DUM has two functions: on one hand, it ensures that all segments have the same size; on the other hand, it accelerates video transmission by bypassing the kernel. LDM delivers the packet without data if the downstream has the cached segment. The simulation is driven by the real YouTube dataset based on CERNET network topology, and the experimental results show that the proposed NVDC outperforms two representative baselines in terms of cache hit ratio, routing hop count, delivery delay and network load.

Acknowledgements This work is supported by Jilin Provincial Social Science Planning Project (Grant No. 2018JD55).

Appendix

The abbreviations frequently used in this paper are listed as follows.

Abbreviations AICDN, AI based mechanism in CDN; ACHR, Average Cache Hit Ratio; ADD, Average Delivery Delay; ANL, Average Network Load; ARHC, Average Routing Hop Count; CC, Constraint Condition; CDN, Content Delivery Networks; CR, Content Router; CS, Content Store; DPDK, Data Plane Development Kit; DIT, Delivery Information Table; DUM, DPDK Usage Module; GRO, Generic Receive Offload; GSO, Generic Segmentation Offload; ICN, Information-Centric Networking; ISP, Internet Service Provider; 6CN, IPv6 Content Networking; HICN, Hybrid ICN; LRU, Least Recently Used; LDM, Light Delivery Module; NVDC, Mechanism in This Paper; ICICN, Network Coding based mechanism in ICN; OS, Origin Server; P2PN, Peer-to-Peer Networking; PEM, Popularity Evaluation Module; SDN, Software-Defined Networking

.

References

1. Zhao H, Ge Y, Liu Q, et al. (2017) P2P lending survey: platforms, recent advances and prospects. ACM Trans Intell Sys Technol 8(6):1–28
2. Anjum N, Karamshuk D, Shikh-Bahaei M, et al. (2017) Survey on peer-assisted content delivery networks. Comput Netw 117:79–95
3. Xylomenos G, Ververidis CN, Siris VA, et al. (2014) A survey of information-centric networking research. IEEE Communications Surveys & Tutorials 16(2):1004–1049
4. HICN. https://wiki.fd.io/view/HICN
5. 6CN. http://6cn.io
6. Akamai. wwww.akamai.com
7. ACM-ICN. https://conferences.sigcomm.org/acm-icn/2019
8. Tang L, Huang Q, Puntambekar A, et al. (2017) Popularity prediction of Facebook videos for higher quality streaming. In: Proc. USENIC annual technical conference, pp 111–123

9. Shen H, Chandler H, Wang H (2018) Toward efficient short-video sharing in the YouTube social network. ACM Trans Internet Technol 18(3):1–25
10. Mawji A, Hassanein H (2011) Efficient content distribution for peer-to-peer overylays on mobile ad hoc networks. J Adv Res 2(3):265–279
11. Matsushita T, Yamanka S, Zhao F (2011) A peer-to-peer content-distribution scheme resilient to key leakage. In: Proc. international workshop on information security applications, pp 121–135
12. Gramatikov S, Jaureguizar F, Cabrera J, et al. (2012) Popularity based distribution schemes for P2P assisted streaming of VoD contents. In: Proc. international conferences on advances in multimedia, pp 14–19
13. You W, Mathieu B, Simon G (2013) Exploiting end-users caching capacities to improve content-centric networking delivery. In: Proc. international conference on P2P, parallel, grid, cloud and internet computing, pp 179–185
14. Zhang X, Wang N, Vassilakis VG, et al. (2015) A distributed in-network caching scheme for p2p-like content chunk delivery. Comput Netw 91:577–592
15. Shehab A, Elhoseny M, Hassanien AE (2017) An efficient scheme for video delivery in wireless networks. In: Proc. quantum computing: an environment for intelligent large scale real application, pp 207–225
16. Ghasemkhani H, Li Y, Moinzadeh K, et al. (2018) Contracting models for P2P content distribution. Production and Operations Managemnt 27(11):1940–1959
17. Garmehi M, Analoui M, Pathan M, et al. (2014) An economic replica replacement mechanism for streaming contnet distribution in hybrid CDN-p2p networks. Comput Commun 52:60–70
18. Ibn-Khedher H, Adb-Elrahman E, Kamal AE, et al. (2017) OPAC: an optimal placement algorithm for virtual CDN. Comput Netw 120:12–27
19. Gussun G (2017) Routing-aware partitioning of the interest address space for server ranking in CDNs. Comput Commun 106:86–99
20. Lai J, Fu Q, Moor2 T (2017) Using SDN and NFV to enhave request rerouting in ISP-CDN collaborations. Comput Netw 113:176–187
21. Li J, Lu Z, Tong Y, et al. (2019) A general AI-defined attention networks for predicting CDN performance. Futur Gener Comput Syst 100:759–769
22. Baake P, Sudaric S (2019) Net neutrality and CDN intermediation. Inf Econ Policy 46:55–67
23. Haw R, Hong CS (2012) A seamless content delivery scheme for flow mobility in content centric network. In: Proc. Asia-Pacific symposium on network operations and management, pp 1–5
24. Feng Z, Xu M, Yang Y, et al. (2016) Optimizing content delivery in ICN networks by the supply chain model. In: Proc. IEEE international conference on performance computing and communications, pp 1–8
25. Frangoudis PA, Polyzos GC, Rubino G (2016) Relay-based multipoint content delivery for wireless users in an information-centric network. Comput Netw 105:207–223
26. Son J, Kim D, Kang H, et al. (2016) Forwarding strategy on SDN-based content centric network for efficient content delivery. In: Proc. international conference on information networking, pp 220–225
27. Ren Y, Li J, Li L, et al. (2017) Modeling content transfer performance in information-centric networking. Futur Gener Comput Syst 74:12–19
28. Lv J, Wang X, Ren K, et al. (2017) ACO-inspired information-centric networking routing mechanism. Comput Netw 126:200–217
29. Azgin A, Ravindran R, Wang G (2018) Scalable multicast for content delivery in information centric networks. In: Proc. international conference on computing, networking and communications, pp 105–111
30. Bourtsoulatze E, Thomos N, Saltarin J, et al. (2018) Content-aware delivery of scalable video in network coding enabled named data networks. IEEE Transactions on Multimedia 20(6):1561–1575
31. Araujo P, Batista I, Linder N (2019) Testbed for ICN media distribution over LTE radio access networks. Comput Netw 150:70–80
32. Cheng X, Dale C, Liu J (2008) Statistics and social network of YouTube videos. In: Proc. IEEE/ACM international symposium on quality of services, pp 229–238
33. DPDK. https://www.dpdk.org
34. CERNET. http://www.topology-zoo.org