Peer-to-Peer Networking and Applications (2021) 14:794–806
https://doi.org/10.1007/s12083-020-01024-y

Light forwarding based optimal CCN content delivery: a case study in metropolitan area network

Songzhu Zhang[1] · Xingwei Wang[1,2] · Jianhui Lv[3] · Min Huang[4]

Received: 22 July 2019 / Accepted: 28 October 2020 / Published online: 4 January 2021

Abstract
Content Centric Networking (CCN) has been introduced in recent years to effectively deliver the named content objects rather than relying on the traditional host-to-host communication model, in which a significant and non-negligible characteristic of CCN is in-network caching. In this paper, we stand on the shoulders of in-network caching to propose the light forwarding based approach in order to optimize CCN content delivery, i.e., the process of data routing irrespective of interest routing. At first, we specially devise an additional packet type which has no data entity at the same time the caching/delivery of data is the granularity of packet. Then, we present a generic content delivery scheme based on packet-level caching and new designed packets; in addition, we make a case study in Metropolitan Area Network (MAN) due to some realistic constraints. Finally, we implement the proposed light forwarding based content delivery scheme and compare it with three existing schemes, and the experimental results reveal that it has more efficient performance optimization, such as network load, delivery delay and energy saving.

Keywords CCN · In-network caching · Content delivery · MAN

1 Introduction

It is well known that Content Centric Networking (CCN) has been investigated around ten years between 2009 when Jacobson published his paper titled "networking named content" [1] in ACM CoNEXT 2019. As can be seen from the past research achievements, interest aggregation and in-network caching brought by Pending Interest Table (PIT) and Content Store (CS) respectively, are considered as two prominent contributions of CCN, which have the guiding and reference significance for the current Internet and even the future Internet [2]. However, they are only used at the phase of interest routing—CCN routing consists of interest routing (content location or interest forwarding) and data routing (content delivery) [3]. To be specific, in terms of interest aggregation, when the multiple repeated interests are requested within a certain period of time, the subsequent same interest requests are aggregated at PIT rather than being continuously forwarded. Therefore, upon the first interest request is answered, it means that all interest requests are satisfied. In terms of in-network caching, if CS of some intermediate node holds the requested content, the interest request has the probability to be satisfied in advance. In other words, the design of cache-aware interest routing is well-received, which accelerates content retrieval. Anyway, we maintain the opinion that, the usage of interest aggregation may be only at the phase of interest routing, while that of in-network caching would be certainly developed at the phase of data routing. On such matter, this paper will study the development of in-network caching for

✉ Jianhui Lv
lvjianhui2012@163.com

Songzhu Zhang
zhangsongzhu@neuq.edu.cn

Xingwei Wang
wangxw@mail.neu.edu.cn

Min Huang
mhuang@mail.neu.edu.cn

1 College of Computer Science and Engineering, Northeastern University, Shenyang, 110819, China

2 State Key Laboratory of Synthetical Automation for Process Industries, Shenyang, 110819, China

3 International Graduate School at Shenzhen, Tsinghua University, Shenzhen 518057, China

4 College of Information Science and Engineering, Northeastern University, Shenyang 110819, China

content delivery but not pay attention to the research of on-path or off-path. In the following, we use term "content delivery" to replace term "data routing".

In addition, during the process of content delivery, the multiple same contents are sent repeatedly, which greatly aggravates network traffic overhead and thus decreases delivery performance of network and Quality of Experience (QoE) of user [4]. As a matter of fact, similar to interest aggregation, the multiple repeated contents can be aggregated within a certain period of time. A feasible method is simply described as follows. If the downstream has the delivering content, the current node only forwards the corresponding content name to its downstream, in which the content is not continuously delivered, that is to say, the forwarding is light. With respect to this method, this paper will present its detailed design standing on the shoulders of in-network caching.

Although there have been many optimization proposals for CCN content delivery, to our best knowledge, the proposed light forwarding based scheme is considerably novel. In particular, it is worth pointing out that the current proposals are usually content-oriented or chunk-oriented, that is, the granularity of caching/delivery is the whole content or the chunk with a certain size. As a matter of fact, it is unscientific in the practice network environment, because the caching/delivery of content-level or chunk-level goes against the share of hot segments and the improvement of cache utilization. For example, if the multiple users want to see hot segments of a video, the upstream has to deliver the repeated and whole video even though these users are not interested in the other parts of the video. Therefore, the more fine-grained caching/delivery, i.e., packet-level based strategy should be considered in order to reach the genuine effect of light forwarding. As a conclusion, the light forwarding has two implications, i.e., identity-based delivery as many as possible and packet-level based caching and delivery. Given the above consideration, the major contributions of this paper are summarized as two points.

- We stand on the shoulders of in-network caching and present a generic content delivery scheme based on the granularity of packet-level to support all application scenarios, where an additional data packet type which only supports the forwarding of index is devised.
- Given some realistic constraints (e.g., information consistency), we make a closed-loop case study in Metropolitan Area Network (MAN) including the design of replacement notification based solution.

The remainder of this paper is organized as follows. Section 2 reviews the related work. In Section 3, the generic content delivery scheme based on light forwarding is presented. Section 4 makes a case study in MAN. The experimental results are reported and analyzed in Section 5. Finally, Section 6 concludes and discusses this paper.

2 Related work

In recent years, a number of CCN content delivery schemes have been proposed. In this section, we only review and summarize some highly relevant work in order to save space and avoid redundancy.

In particular, the fundamental content delivery prototype was introduced in [5], where the data was carried to interest requester along the reverse and symmetric path without any performance optimization. Although [6] built a novel secure content delivery framework to enable users to receive the encrypted content cached at a nearby router by a popular symmetric key encryption algorithm, it only paid attention to the security rather than the efficiency of content delivery. In addition, an emergent message delivery mechanism for the special telephone-based call situation by leveraging multicast was proposed in [7], similar to [6], it did not consider the concrete optimization for content delivery.

In [8], authors characterised a multi-path content transfer strategy, which divided the content into some chunks and cached them at the distributed intermediate nodes. However, [8] was subject to two inevitable assumptions where each interest requester had the similar number of paths connecting to the cache nodes and each cache node had the single path to server. Furthermore, regarding the multi-path delivery, [9, 10] simulated the behavior of ants and devised two related content delivery schemes. However, the delivery based on Ant Colony Optimization (ACO) had to undergo several iterations, which was confronted with the enormous challenges in terms of the real-time services.

In [11–13], the network coding based content delivery methods were proposed to optimize network throughput and efficiency. In spite of this, it was unavoidable that the network coding brought high time complexity and greatly increased computation overhead. In [14], inspired by product delivery which belonged to the economic field, authors leveraged the inventory model of supply chain management in logistics to formulate the process of content delivery. However, [14] made a naive assumption where all interest requests were routed to the nearest content providers, which could not be certainly guaranteed in the actual forwarding environment. In [15], authors exploited Markovian queuing system theory and computed the queuing delays of data packets, in which the delivery order of data packets depended on their corresponding queuing delays. Although the queuing theory based delivery strategy optimized the overall performance, the total delivery traffic did not decrease and thus the network still faced the heavy load. In [16], a scalable multicast scheme was proposed to deliver the content to the multiple interest requesters. However, [16] was based on the push-mode rather than pull-mode, which was beyond the principle of CCN.

Although the above mentioned proposals optimized the performance of delivery, they did not effectively decrease the total delivery traffic. In fact, if the downstream has the delivering content due to the ability of in-network caching, it is unnecessary that the current node does the repeated content delivery, which is our basic idea. In addition, completely different from the previous work, the granularity of caching/delivery in this paper is packet-level instead of content-level or chunk level.

3 Generic delivery design

3.1 Problem description

For the sake of understanding, the abbreviations frequently used in this paper are listed in Table 1. In this paper, we regard the intermediate node with the cache ability as Content Router (CR). Suppose that there are n CRs in the given network topology, and they are denoted by $CR_1, CR_2, \cdots, CR_n$. Consider a scenario in Fig. 1a which consists of five CRs, delivering the content <video-X,data> with 20M to user1 from CR_1 and to user2 from CR_2 respectively. At the initial, CR_3 and CR_5 have <video-X,data> while CR_1, CR_2 and CR_4 have no.

In terms of the traditional delivery scheme, CR_1 and CR_2 deliver <video-X,data> to user1 and user2 respectively via CR_3, CR_4 and CR_5 in turn, that is, the corresponding data of video-X is delivered without omission, as shown in Fig. 1b. In fact, delivering <video-X,data> from CR_1 and CR_2 to CR_3, from CR_3 to CR_4 and from CR_4 to CR_5 is unnecessary if CR_1 and CR_2 know that CR_5 holds <video-X, data>. It is obvious that the network pressure of delivering heavy traffic is alleviated. To be specific, instead of delivering <video-X,data>, both CR_1 and CR_2 deliver <video-X> to CR_5 and then CR_5 loads the corresponding data to deliver <video-X,data> to user1 and user2, as shown in Fig. 1c.

Table 1 Abbreviations in alphabetical order

Number	Abbreviation	Full name
1	ACO	Ant Colony Optimization
2	BRAS	Broadband Remote Access Server
3	CCN	Content Centric Networking
4	CDN	Content Delivery Network
5	CR	Content Router
6	CS	Content Store
7	FIB	Forwarding Information Base
8	LRU	Least Recently Used
9	MAN	Metropolitan Area Network
10	NPT	New Packet Type
11	PIT	Pending Interest Table
12	QoE	Quality of Experience
13	KP01	0-1 Knapsack Problem

It is a fact that users are usually interested in the hot segments of a video. Therefore, similar to Content Delivery Network (CDN), CCN content delivery scheme should also support the mode of Video-on-Demand (VOD). However, different from CDN, it is not likely to design the big cache size for CR of CCN. Under such context, the caching of content-level or chunk-level cannot significantly enhance the efficiency of delivery and the utilization of cache. Given this, the packet-level based caching/delivery is explored.

Integrating with the above descriptions, the fundamental idea of light-forwarding based content delivery is to check whether the downstream has the delivering packet rather than content or chunk: if yes, deliver packet without data (i.e., the light packet); otherwise, deliver packet with data (e.g., 1500B).

Certainly, the existence of problem may be criticized by some readers. For example, they argue that, it is not likely for CR_1 to deliver <video-X,packet-Y,data> in order to satisfy the request of user1, because CR_5 has <video-X,packet-Y,data> and it should satisfy video-X request of user1 in advance. However, in the practice network environment, our problem description is existed and reasonable, and two illustrations are listed as follows. (1) When video-X request arrives at CR_5, CR_5 has no <video-X,data>; but before CR_1 delivers <video-X,packet-Y>, it is likely that, the other CRs deliver <video-X,packet-Y,data> to the other users via CR_5 and CR_5 caches <video-X,packet-Y,data>. (2) CR_5 is regarded as the nearest content provider. However, it is very difficult and even impossible that each user can find the correspondingly nearest content provider. For example, CR_5 is not the unique selection to forward video-X request to CR_1.

3.2 Solution presentation

In this section, we first introduce two new-style packets to replace the traditional data packet during the process of content delivery. Then, based on packet-level caching/delivery, we design a solution to optimize content delivery. Finally, given some realistic constraints in mesh network topology, we discuss the necessity of research in the special scenario.

3.2.1 New-style packets

The traditional data packet in CCN has two parts, i.e., content name (header) and data (payload or content), in which content name is used to identify the corresponding

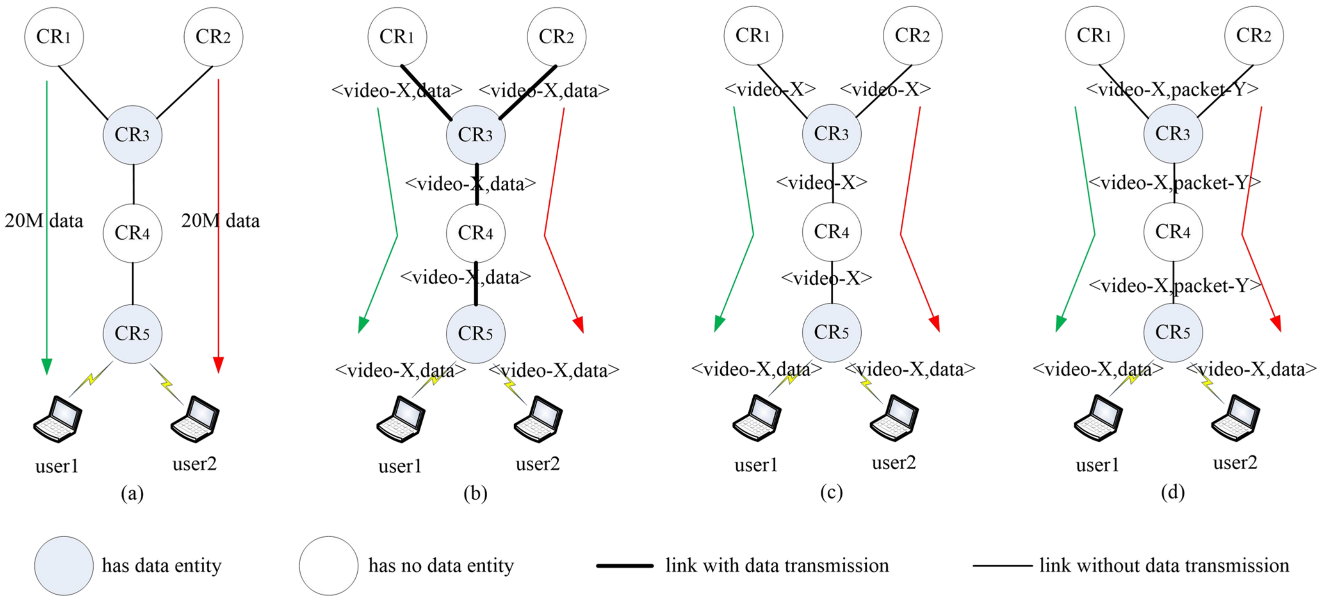


Fig. 1 An example for problem description

data. As the above mentioned, the traditional data packet cannot satisfy the optimal content delivery, and thus we need to devise the additional packets for optimizing content delivery. In Fig. 2, we devise two new-style packets. The first New Packet Type (NPT1) consists of content name, IDentity (ID) order and data. The second NPT (NPT2) only has content name and ID order but no data. In particular, a content name corresponds to a number of ID orders and each small data (e.g.,1500B) is identified by an unique ID order.

For example, the required content size is 2M and the small data size is 1500B, and we have that the number of ID orders is 1399 ($\lceil 2*1024*1024/1500 \rceil$). In terms of this example, the size of the traditional data packet is 2M (i.e., content-level or chunk-level), while the size of the new-style packet (i.e., NPT1) is 1500B (i.e., packet-level). In particular, the caching of packet-level can not only enhance the efficiency of content delivery by increasing the utilization of cache but also facilitate the compatibility with the current IP networks.

3.2.2 Delivery flow

Suppose that the content name is NTXXI21, if the downstream has the information of <NTXXI21,20>, instead of delivering <NTXXI21,20,1500B>, the current CR only delivers <NTXXI21,20> to its downstream. Based on the two new-style packets i.e., NPT1 and NPT2 and packet-level caching/delivery, we present a generic content delivery scheme, as follows.

Algorithm 1 Generic content delivery scheme.

```
if CR receives NPT1, then
  if the downstream has NPT1, then
    //NPT1 is not cached at CR;
    Delivers NPT2;
  else //the downstream has no NPT1
    if CR has no NPT1, then
      CR caches NPT1;
    end if
    Delivers NPT2;
  end if
else //CR receives NPT2;
  if the downstream has NPT1, then
    //NPT1 is not cached at CR;
    Delivers NPT2;
  else //the downstream has no NPT1 while
      //CR certainly has NPT1;
    Recover NPT1 from NPT2;
    Deliver NPT1;
  end if
end if
```

In Algorithm 1, the judgement of information consistency is considerably important, that is, how to address the

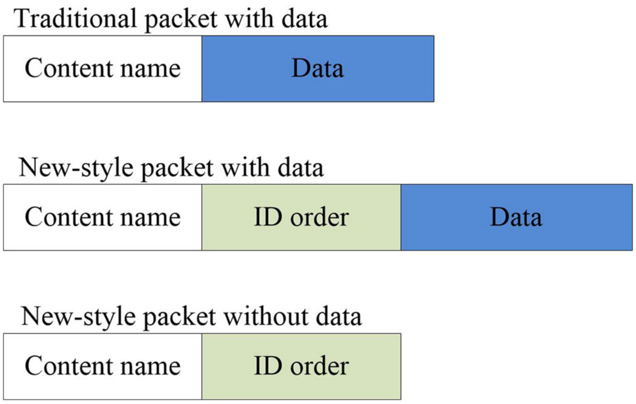


Fig. 2 Packet format including traditional data packet and two new-style packets

issue with respect to “the downstream has (no) NPT1”. The definition of information consistency is described as follows.

Definition 1 Consider that CR judges the downstream has (no) NPT1: if the judgement is correct, it is called information consistency; otherwise, it is called information inconsistency.

3.2.3 Necessary analysis

According to Algorithm 1 and Definition 1, we observe that the guarantee of information consistency is the fundamental prerequisite. However, the nature guarantee without introducing the auxiliary mechanism is very difficult, especially in mesh network topology. For example in Fig. 1, two special situations with respect to the judgement of <NTXXI21,20,1500B> are analyzed, where a certain period of time is 2 min.

- (1) At 0.5 min, CR_1 judges that CR_3 has <NTXXI21,20,1500B>. However, due to CR_2 delivers some new contents to CR_3 continuously, which causes the frequent replacement so that <NTXXI21,20,1500B> has to be removed from CR_3 at 1 min. At 1.2 min, CR_1 is sure of that CR_3 has <NTXXI21,20,1500B>, which causes information inconsistency. If CR_1 continues to deliver <NTXXI21,20> to CR_3, CR_3 has no ability to do the follow-up delivery.
- (2) Within 1 min, CR_1 judges that CR_3 has no <NTXXI21,20,1500B>. However, due to CR_2 delivers <NTXXI21,20,1500B> to CR_3 and <NTXXI21,20,1500B> is cached at CR_3 within 1.2 min, which causes information inconsistency. If CR_1 delivers <NTXXI21,20,1500B> to CR_3, which increases the redundant overhead. In other words, it only requires CR_1 to deliver <NTXXI21,20> to CR_3.

On the basis of the above example analysis, we have the following theorem in order to illustrate the guarantee difficulty of information consistency in mesh network topology due to its NP-hardness.

Theorem 1 *Considering mesh network topology, the guarantee of information consistency is an NP-hard problem. The related proof is found in* Appendix.

4 A case study in MAN

According to Theorem 1, we know that the guarantee of information consistency in mesh network topology is an NP-hard problem. In order to verify the feasibility and effectiveness of the proposed generic content delivery scheme, we make a case study in MAN scenario, in which the issue of information consistency can be easily addressed.

The abstract MAN scenario is depicted in Fig. 3, which involves four layers, i.e., m servers, a few ordinary router/switches, n Broadband Remote Access Servers (BRASs) and a large amount of users. In particular, these ordinary routers/switches are only used to forward packets; each BRAS network segment is deployed for one CR and several users; these different BRASs do not overlap mutually.

4.1 Preliminary

Server is equipped with Forwarding Information Base (FIB) which consists of four fields, i.e., content name, ID order, timer and BRASip. Since the physical topology of MAN is relatively simple, at initial, BRASip can be configured in the server. An entry <NTXXI21,20,5 min,10.71.149.39> in FIB means that the server delivers <NTXXI21,20> with data to the network segment of 10.71.149.39 and the survival time of the delivery information is 5 min.

CR is equipped with CS which consists of five fields, i.e., content name, ID order, data, timer and Serverip. An entry <NTXXI21,20,1500B, 5 min, 16.23.44.12> in CS means that the CR receives <NTXXI21,20,1500B> from 16.23.44.12 and the survival time of data is 5 min. Although CR also has PIT and FIB, we do not further consider this matter since both downstream and upstream of CR have no other CRs.

With respect to timer, it has a fixed value, denoted by T. That is to say, when a new entry in both FIB and CS is generated, the corresponding timer is set as T. In addition, some special illustrations for timer are summarized as follows. (1) With a certain period of time,

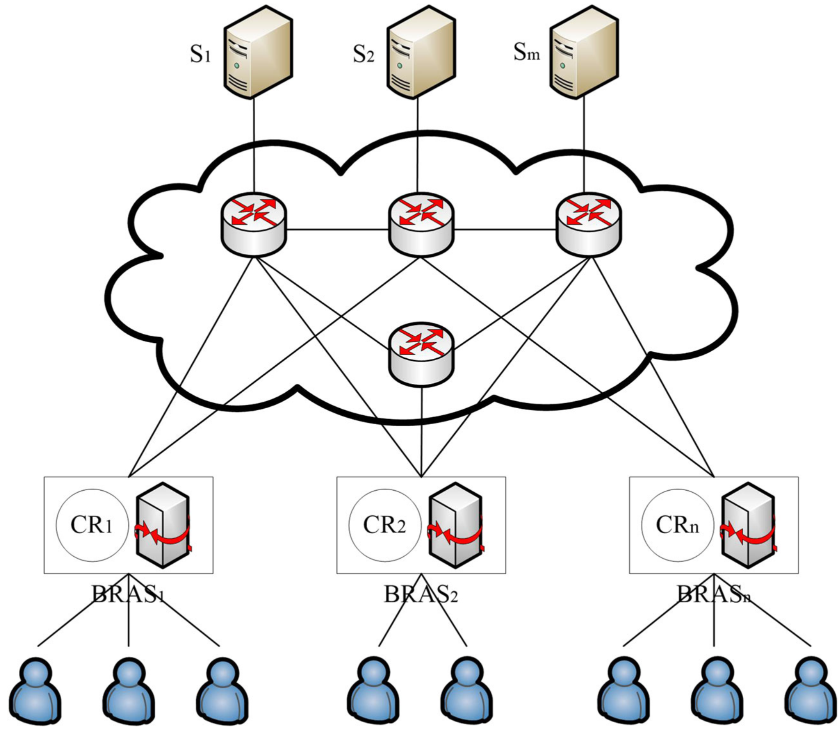


Fig. 3 An abstract MAN scenario with four layers

upon server delivers the multiple repeated information to the downstream, the timer in FIB is reset as T. (2) With a certain period of time, upon CR receives the multiple recreated information from the upstream, the timer in CS is reset as T. (3) As time goes on, if the timer of some entries in FIB or CS becomes 0, these entries are deleted automatically. (4) In order to avoid the phenomena of timer inaccuracy, when an entry in CS is generated, its timer should be $T + \Delta$, here Δ is a very small value (e.g., $20us$).

4.2 Replacement notification

The information inconsistency is caused by the packet replacement which frequently happens at the downstream CR. Then, the core idea of replacement notification based solution is to make replacement transparent with the following three major steps. (1) CR receives an NPT1 while its CS has no enough space to store the NPT1. (2) One NPT1 is replaced by Least Recently Used (LRU) and the new arrival NPT1 is cached. (3) CR generates a notification message and sends it to all possible upstream servers. Here, we note that the "Serverip" field of CS has to be improved in order to record the multiple same NPT1s but from different servers (e.g.,<NTXXI21,20,1500B,3 min,Server1ip/Server2ip>). In particular, the format of notification message can have three parts, i.e., operation, content name and ID order, and the special "operation" is used to differentiate from the regular NPT2, where the concrete meaning of "operation" is diverse, such as delete (001), source quench (010), source recovery (100), etc.

For example, <NTXXI21,20,1500B,3 min,Server1ip/Server2ip> is removed from CR_1. At first, CR_1 sends <001,NTXXI21,20> to both server1 and server2. Then, server1 and server2 delete <NTXXI21,20,2 min,BRAS1> and <NTXXI21,20,3 min,BRAS1> from their FIBs respectively. The pseudo-code of replacement notification based solution is described in Algorithm 2.

Algorithm 2 Replacement notification based solution.

CR receives $NPT1_i$;
//CR checks whether CS has enough space;
if CR has ability to cache $NPT1_i$, **then**
  Cache $NPT1_i$;
else
  $NPT1_j$ is replaced by LRU;
  Cache $NPT1_i$;
  Generate a notification message on $NPT1_j$;
  Send the notification message to all possible servers;
end if
begin in parallel
  $Server_i$ receives the notification message;
  $Server_i$ deletes the entry on $NPT1_j$ from FIB;
end in parallel

In terms of the proposed Algorithm 2, it is used for solving the information consistency from the perspective of network by sending notification message, which does not affect the network performance. In fact, when and only when the superabundant notification messages are generated, the network performance would only get into the difficulties. To be specific, the redundant messages cause the serious network congestion, which further increase network delay and bring into the bad QoE. However, it is usually impossible with respect to generating the superabundant notification messages suddenly, this is because the frequency of content replacement which causes to generate the notification messages is not so high in practice. Here, we emphasize that this paper does not design the specialized method to reduce the frequency of content replacement. In spite of this, the inherent content replacement, i.e., first in, first out can satisfy the proposed application scenario. Based on such analysis, the information consistency can be guaranteed in a bounded performance.

5 Experimental results

5.1 Simulation settings

The proposed Light Forwarding based Content Delivery scheme, LFCD for short, is simulated over NS3 [17] based on C++, runing on a personal computer with Intel(R) Xeon e5-2680, 2.39 GHZ CPU, 7.99 GB RAM over windows 8 system. The performance is evaluated by considering three metrics, i.e., network load, delivery delay and energy efficiency. Among them, (1) the network load is defined as the ratio of the transmitted traffic and the maximum transmission traffic within a certain period of time; (2) the delivery delay is defined as the time-slot between the point-in-time when a requested content (e.g., 20M) is sent from content provider and that when interest requester receives the sent content; (3) the energy efficiency is linearly proportional to the transmitted traffic and its definition can refer to [18]. In addition, to verify the effectiveness of LFCD, we compare it with three the-state-of-the-art and representative schemes, i.e., those in [10, 13] and [15] which are shortened to S10, S13 and S15 respectively.

5.2 Dataset and network environment

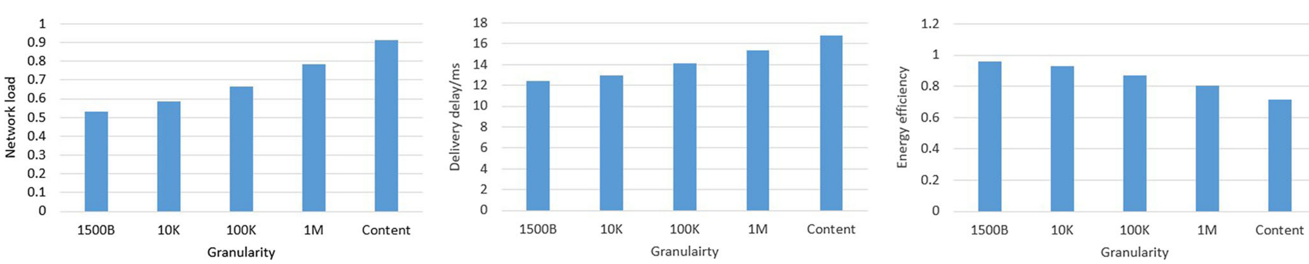
The simulation is driven based on the real dataset, of which the trace collection comes from a campus network measurement on YouTube short videos [19], and some detailed illustrations on YouTube dataset are presented as follows. (1) Send-end: there are 1563 servers and 2600 contents, in which the sizes of 80% contents are smaller than or equal to 12M. (2) Receiver-end: there are 89 network segments and 2377 users, including 600000 interest requests. (3) Data distribution: the arrival time with respect to interest requests follows the Poisson distribution; the number of content deliveries from servers follows the power-law distribution; the number of interest requests from network segments follows the power-law distribution.

According to the above statements on YouTube dataset, the network environment is set as follows. (1) The physical network topology is the same as Fig. 3 which is also used for simulation, consisting of 1563 servers, 89 network segments and 2377 users. (2) All 1563 servers provide 2600 contents, and the providing law follows the power-law distribution. (3) Each network segment is deployed for one CR, and each CR has $\lceil 2377/89 \rceil$ users. (4) All 2377 users send 600000 interest requests, and the sending law follows the power-law distribution. (5) The cache size of CR is 40M the lifecycle of cache item is variable (i.e., $T = 2$ min, 3 min, 4 min, 5 min and 6 min). In other words, the features of Youtube dataset are reflected to Fig. 3 for simulation.

5.3 Scheme verification

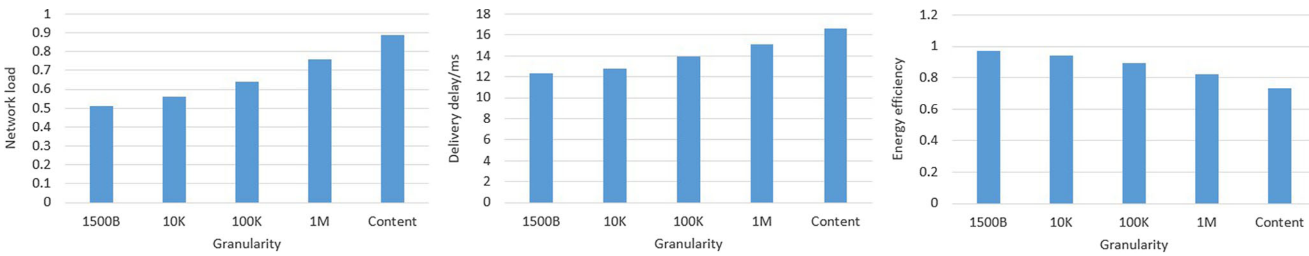
In this section, we verify the feasibility of LFCD by randomly selecting content deliveries for 100 times. With respect to three evaluation metrics, the experimental results on mean are reported in Figs. 4, 5, 6, 7, and 8 (different settings of T) and Figs. 9, 10, 11, 12 and 13 (different size settings of transmission unit, i.e., packet-level, chunk-level and content-level).

According to Figs. 4a–8a, we observe that, the content-level based caching/delivery causes the most serious network load while the packet-level based caching/delivery



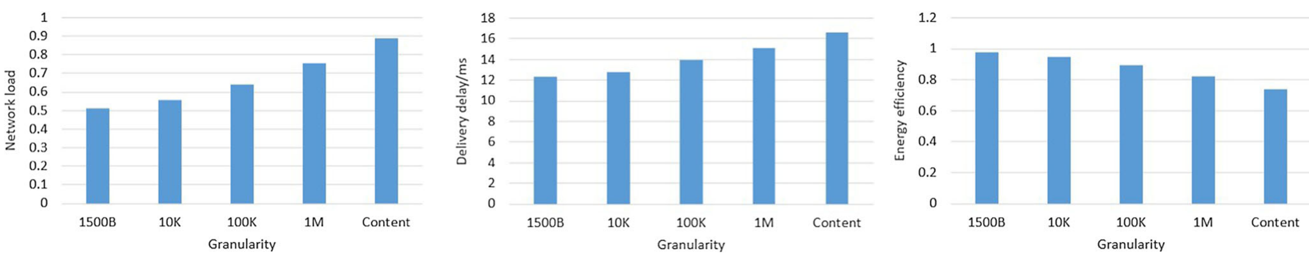
(a) Network load (b) Delivery delay (c) Energy efficiency

Fig. 4 The experimental results on network load, delivery delay and energy efficiency when $T = 2$ min



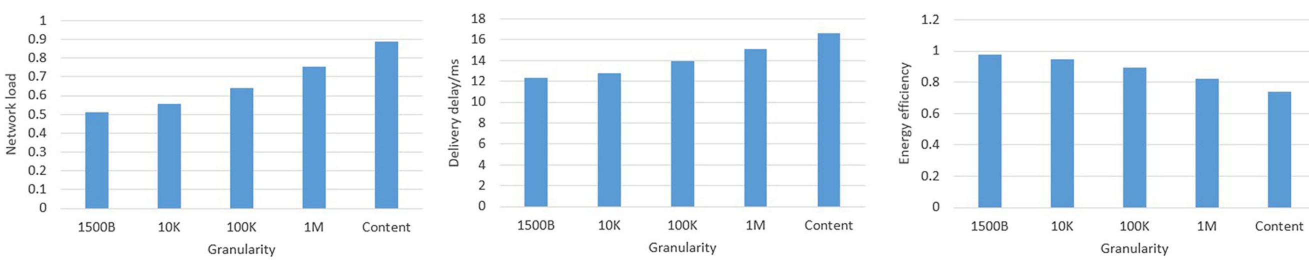
(a) Network load (b) Delivery delay (c) Energy efficiency

Fig. 5 The experimental results on network load, delivery delay and energy efficiency when $T = 3$ min



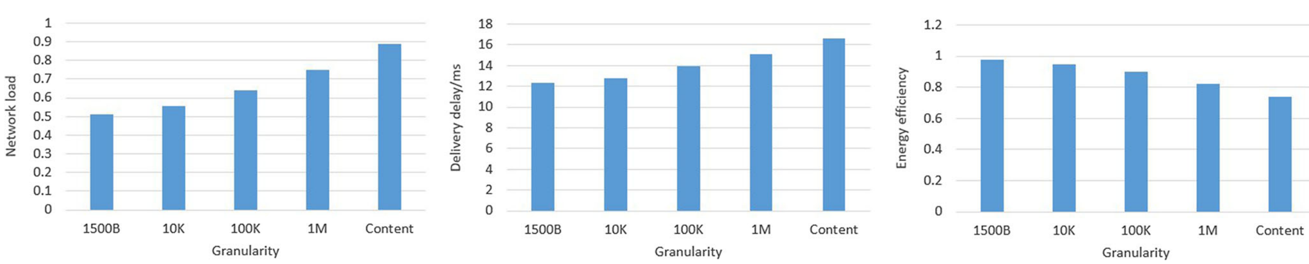
(a) Network load (b) Delivery delay (c) Energy efficiency

Fig. 6 The experimental results on network load, delivery delay and energy efficiency when $T = 4$ min



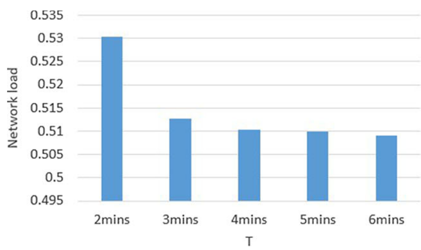
(a) Network load (b) Delivery delay (c) Energy efficiency

Fig. 7 The experimental results on network load, delivery delay and energy efficiency when $T = 5$ min

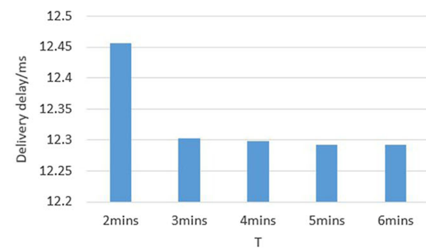


(a) Network load (b) Delivery delay (c) Energy efficiency

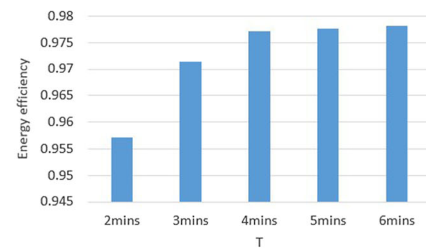
Fig. 8 The experimental results on network load, delivery delay and energy efficiency when $T = 6$ min



(a) Network load

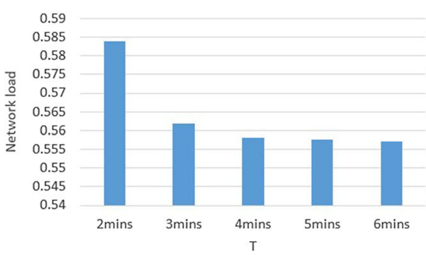


(b) Delivery delay

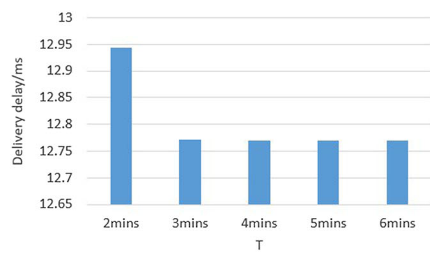


(c) Energy efficiency

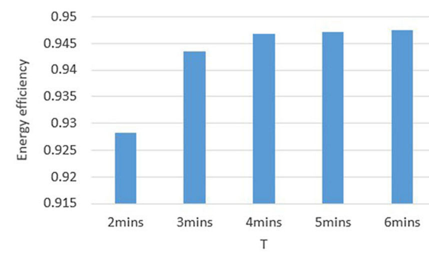
Fig. 9 The experimental results on network load, delivery delay and energy efficiency with packet-level (1500B)



(a) Network load

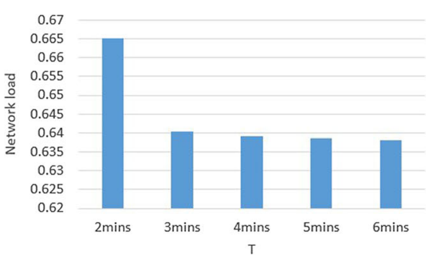


(b) Delivery delay

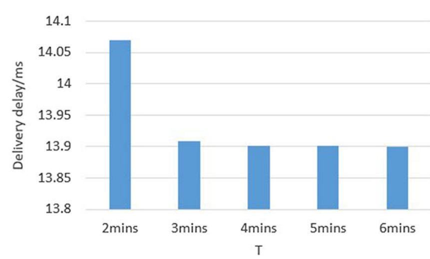


(c) Energy efficiency

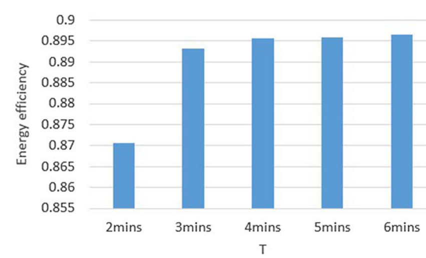
Fig. 10 The experimental results on network load, delivery delay and energy efficiency with chunk-level (10K)



(a) Network load

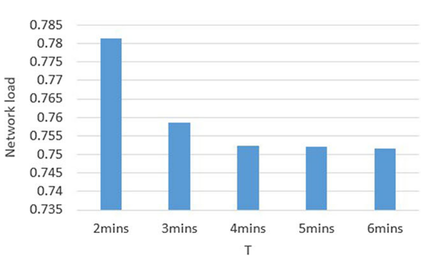


(b) Delivery delay

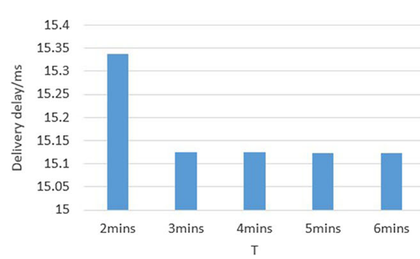


(c) Energy efficiency

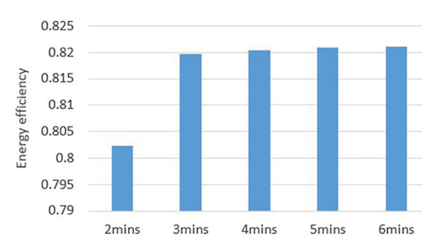
Fig. 11 The experimental results on network load, delivery delay and energy efficiency with chunk-level (100K)



(a) Network load

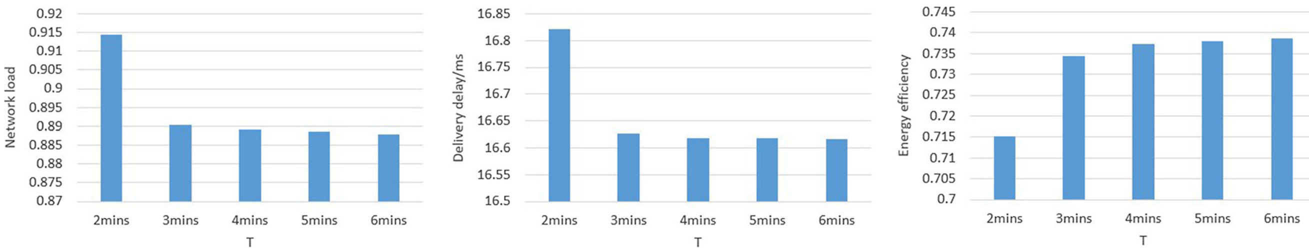


(b) Delivery delay



(c) Energy efficiency

Fig. 12 The experimental results on network load, delivery delay and energy efficiency with chunk-level (1M)



(a) Network load (b) Delivery delay (c) Energy efficiency

Fig. 13 The experimental results on network load, delivery delay and energy efficiency with content-level

can reach the genuine effect of light forwarding with the smallest network load. According to Figs. 4b–8b, we observe that, the delivery delay based on packet-level is the most acceptable (only around 12.3ms); this is because, oriented-chunk or -content caching/delivery easily causes the redundant data transmission, which has to consume much more time. According to Figs. 4c–8c, we observe that, the size of transmission unit with 1500B has the best energy-efficient effect; this is because, CR usually caches the hot segments and it only needs to pull a small quantity of and new segments from servers, which saves a mass of traffic transmission.

According to Figs. 9–13, we observe that the performance of LFCD has been continually increasing. However, after T reaches 3 min, network load, delivery delay and energy efficiency show no significant change. This indicates that 3 min cache period and 40M cache space can satisfy all interest requests for hot segments.

In summary, when the transmission unit is set as 1500B and the cache period is set as 3 min, LFCD has the best performance. It suggests that the packet-level based light caching/delivery is feasible and the obtained performance is acceptable.

5.4 Comparison analysis

The network load for LFCD, S10, S13 and S15 is reported in Fig. 14. We observe that LFCD faces the smallest network load, followed by S13, S10 and S15. Some reasons are listed as follows. (1) Different from S10, S13 and S15, LFCD uses the light forwarding strategy to deliver a large amount of NPT2s instead of NPT1s, which avoids transmitting the redundant data and thus the corresponding network load is the lightest. (2) According to the experimental results from Fig. 4a–c, we know the content-level based caching/delivery has the most serious network load. In fact, among S10, S13 and S15, different from S10 and S13 both with the chunk-level based caching/delivery, the caching/delivery granularity of S15 is the content-level. Thus, S15 faces more serious network load than S10 and S13. (3) Although S10 and S13 are subject to the chunk-level based caching/delivery, S13 does network coding and has more fine-grained caching/delivery than S10; in addition, when S10 delivers the whole content, a lot of ants are generated and sent to carry their corresponding chunks. It is obvious that the generation of ants increases network load. With the two aspects consideration, S13 has smaller network load than S10.

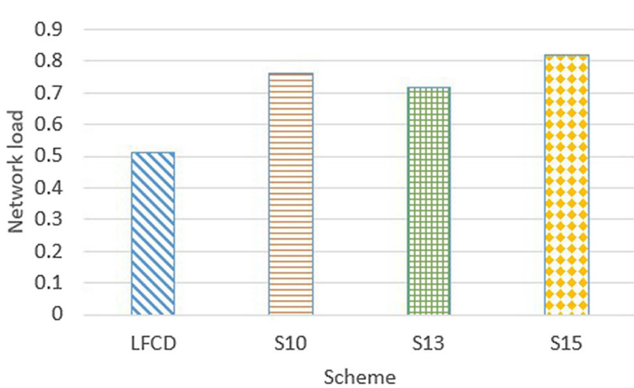


Fig. 14 Comparison results on network load

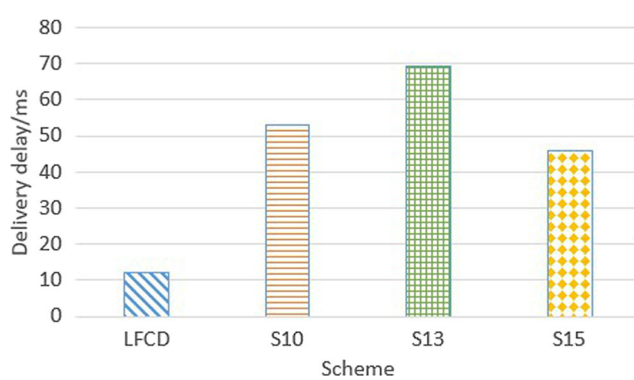


Fig. 15 Comparison results on delivery delay

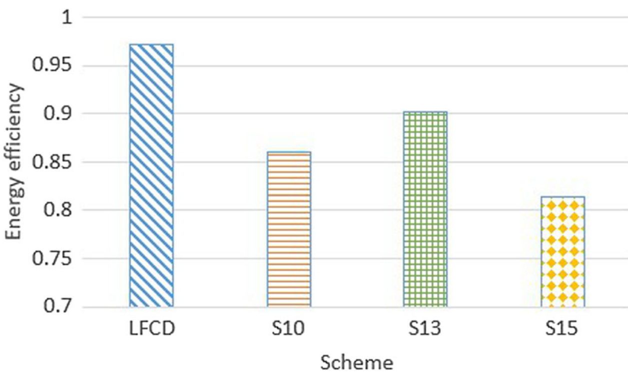


Fig. 16 Comparison results on energy efficiency

The delivery delay for LFCD, S10, S13 and S15 is reported in Fig. 15. We observe that LFCD has the smallest delivery delay, this is because, CR caches the overwhelming majority of hot segments and servers only need to deliver the corresponding NTP2s (the transmission time of NPT2 is far smaller than that of NPT1). For the other three schemes, S15 has the smallest delivery delay, followed by S10 and S13, because S15 only uses Markovian queuing system theory instead of S10 with ACO and S15 with network coding. Although ACO of S10 and network coding of S13 bring high time complexity and large computation overhead, ACO in MAN scenario does not undergo the multiple iterations and thus has smaller delivery delay than S13.

The energy efficiency for LFCD, S10, S13 and S15 is reported in Fig. 16. We observe that LFCD has the highest energy efficiency, followed by S13, S10 and S15. Some major illustrations are summarized as follows. (1) LFCD delivers the minimum network traffic due to the introduction of light forwarding and has the most energy-efficient effect. (2) S15 still exploits Markovian queuing system theory to compute the queuing delays of data packets, that is to say, servers are constantly in force, which needs too much energy consumption. (3) S10 faces more traffic than S13 due to the generation of ants and thus has worse energy-efficient effect than S13. In fact, from the perspective of traffic (see Fig. 14), with a certain of period time, LFCD transmits the smallest traffic, followed by S13, S10 and S15. Since the energy efficiency is linearly proportional to the traffic, it is obvious that LFCD has the most energy-efficient effect, followed by S13, S10 and S15.

6 Conclusion and discussion

CCN is an evolving networking paradigm, which allows the content to be cached at the intermediate router due to the inherent ability of in-network caching. However, the traditional CCN usually delivers the redundant contents during the process of content delivery, which not only increases network overhead but also reduces user's QoE. To such end, this paper proposes a light forwarding based generic approach to optimize content delivery; in particular, the light forwarding has two implications: identity-based delivery as many as possible and packet-level based caching/delivery. Since the information consistency is very hard to be addressed in mesh network topology, we make a specific case study in MAN and present the corresponding solution. To verify the effectiveness and efficiency of the proposed scheme, we compare it with three the-state-of-the-art representatives by considering network load, delivery delay and energy efficiency.

Although the light forwarding based idea is very novel and interesting, it remains an issue which is considerably difficult to be addressed. To be specific, the proposed Algorithm 1 can be regarded as a nice solution for all kinds of topologies such as ring network topology, bus network topology, star network topology, tree network topology (a typical case MAN) and mesh network topology when and only when the information consistency can be guaranteed. Although the combination of Algorithms 1 and 2 can solve the tree network topology, the other kinds of network topologies cannot be solved well. As matter of fact, only tree network topology and mesh network topology are two common and significant ones in the future network development, thus we plan to study the solution for overcoming the information consistency about mesh network topology in the next phase. In addition, we still maintain that the in-network caching is extremely valuable while the cache size should be designed as small as possible due to the limitation of price. Regarding this, we plan to study the trade-off between cache size and benefit.

Acknowledgments This work is supported by the LiaoNing Revitalization Talents Program under Grant No. XLYC1902010, the Major International (Regional) Joint Research Project of NSFC under Grant No. 71620107003, and the National Natural Science Foundation of China under Grant No. 61872073.

Appendix

The proof of Theorem 1.

(1) Suppose that CS of CR_i has q_i entries. If the guarantee of information consistency is correct, it means that the replacement situation of each entry is transparent. On this basis, consider mesh network topology, the essential condition on information consistency is that each CR knows all replacement situations. Let N denote the total combination number of all replacement situations, and we have

$$N = \prod_{i=1}^{n} q_i \tag{1}$$

(2) Next, we consider the 0–1 Knapsack Problem (KP01) [20], which is described as follows. Given n items that can be regarded as n CR in the network environment, where item i (i.e., CR_i in the network environment) owns weight w_i and profit p_i, and the knapsack has a fixed capacity M. The ultimate goal of KP01 is to obtain the maximal profit by loading some possible items into the knapsack. Mathematically, we have

$$\begin{cases} Maximize & \sum_{i=1}^{n} p_i x_i \\ s.t. & \sum_{i=1}^{n} w_i x_i \le M \end{cases} \tag{2}$$

where $x_i \in \{0, 1\}$: $x_i = 0$ means that item i is loaded into the knapsack; otherwise, no. Suppose that the optimal solution of KP01 consists of k items, and the total combinations number of these k items is defined as follows.

$$N' = \binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{3}$$

(3) Reviewing equation (1), in the practical network environment, each CS has a amount of entries, i.e., $q_i > n$ is always satisfied. Furthermore, we have

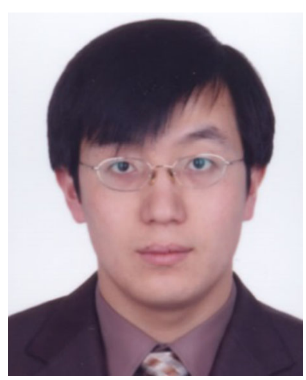
$$N = \prod_{i=1}^{n} q_i > \prod_{i=1}^{n} n > n! > \frac{n!}{k!(n-k)!} = N' \tag{4}$$

which indicates a reduce relationship between the proposed Theorem 1 and KP01. To be specific, the solution space of Theorem 1 can fully cover that of KP01, that is to say, Theorem 1 can solve KP01. As we know, KP01 is an NP-hard problem, thus the problem involved in Theorem 1 is also NP-hard. To sum up (1), (2) and (3), Theorem 1 is proved.

References

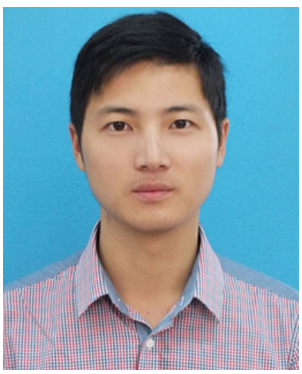
1. Jacobson V, Smetters DK, Thornton JD et al (2009) In: Proceedings of the ACM CoNEXT, pp 1–12
2. Pan J, Paul S, Jain R (2011) A survery of the research on future Internet architecture. IEEE Commun Mag 49(7):26–36
3. Ahlgren B, Dannewitz C, Imbrenda C et al (2012) A survery of information-centric networking. IEEE Commun Mag 50(7):26–36
4. Chen Y, Wu K, Zhang Q (2015) From QoS to QoE: a tutorial on video quality assessment. IEEE Commun Surv Tutor 17(2):1126–1165
5. Yi C, Afanasyev A, Moiseenko I et al (2013) A case for stateful forwarding plane. Comput Commun 36:779–791
6. Misra S, Tourani R, Majd NE (2013) Secure content delivery in information-centric networks: design, implementation and analyses. Proceedings of the ACM SIGCOMM workshop on ICN 73–78
7. Hasegawa T, Tara Y, Ryu K et al (2016) Emergency message delivery mechanism in NDN networks. In: Proceedings of the ACM ICN, pp 199–200
8. Udugama A, Palipana S, Goery C (2013) Analytical characterisation of multi-path content delivery in content centric networks. In: Proceedings of the IEEE international conference on future internet communications, pp 1–7
9. Lv J, Wang X, Ren K et al (2017) ACO-inspired information-centric networking routing mechanism. Comput Netw 126:200–217
10. Lv J, Wang X, Huang M (2018) A smart ACO-inspired named data networking forwarding scheme with clustering analysis. Trans Emerg Telecommun Technol 29(3):1–18
11. Lei K, Zhu F, Peng C et al (2015) Network coding for effective NDN content delivery: models, experiments, and applications. In: Proceedings of the international conference on parallel processing, pp 101–110
12. Ding W, Wang J, Lu K et al (2017) Optimal cache management and routing for secure content delivery in information-centric networks with network coding. In: Proceedings of the IEEE international conference on parallel and distributed processing with applications, pp 267–274
13. Bourtsoulatze E, Thomos N, Saltarin J et al (2018) Content-aware delivery of scalable video in network coding enabled named data networks. IEEE Trans Multimed 20(6):1561–1575
14. Feng Z, Xu M, Yang Y et al (2016) Optimizing content delivery in ICN networks by the supply chain model. In: Proceedings of the IEEE international conference on performance computing and communications, pp 1–8
15. Ren Y, Li J, Li L et al (2017) Modeling content transfer performance in information-centric networking. Future Gener Comput Syst 74:12–19
16. Azgin A, Bavindran R, Wang G (2018) Scalable multicast for content delivery in information centric networks. In: International conference on computing networking and communications, pp 105–110
17. NS3, https://www.nsnam.org/
18. Wang X, Lv J, Huang M et al (2018) Energy-efficient ICN routing mechanism with QoS support. Comput Netw 131:35–51
19. YouTube traces from the campus network, http://traces.cs.umass.edu/index.php/Network/Network
20. Lv J, Wang X, Huang M et al (2016) Solving 0–1 knapsack problem by greedy degree and expectation efficiency. Appl Soft Comput 41:94–103

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Songzhu Zhang received the B.S. degree in computer science from the Northeastern University, Qinhuangdao China in 2003 and received the M.S. degree in computer science from the Yanshan University, Qinhuangdao China in 2009. He is currently working toward the Ph.D degree in the Northeastern University, Shenyang, China. His research interests include Information-Centric Networking, Software-Defined Networking, etc.

Xingwei Wang received the B.S., M.S., and Ph.D. degrees in computer science from the North-eastern University, Shenyang, China in 1989, 1992, and 1998 respectively. He is currently a Professor at the College of Computer Science and Engineering, Northeastern University, Shenyang, China. His research interests include cloud computing and future Internet, etc. He has published more than 100 journal articles, books and book chapters, and refereed conference papers. He has received several best paper awards.



Jianhui Lv received B.S. degree in mathematics and applied mathematics from the Jilin Institute of Chemical Technology, Jilin, China in 2012, and M.S. & Ph.D degree in computer science from the Northeastern University, Shenyang, China in 2014 & 2018. He worked at the Network Technology Lab, Central Research Institute, Huawei Technologies Co. Ltd, Shenzhen, China as a senior engineer from Jan. 2018 to Jul. 2019. He is currently an assistant research fellow at the International Graduate School at Shenzhen, Tsinghua University. His research interests include ICN, in-network caching enabled networks, IoT, edge computing, etc. He has published more than 20 journal and conference papers, such as Elsevier COMNET, Elsevier ASOC, Elsevier JNCA, IEEE COML, IEEE INFOCOM, IEEE/ACM IWQoS, etc. He is a member of the IEEE.

Min Huang received the B.S. degree in automatic instrument, the M.S. degree in systems engi-neering, and Ph.D. degree in control theory from the Northeastern University, Shenyang, China in 1990, 1993, and 1999 respectively. She is currently a Professor at the College of Information Science and Engineering, Northeastern University, Shenyang, China. Her research interests include modeling and optimization for logistics and supply chain system, etc. She has published more than 100 journal articles, books, and refereed conference papers.