Research Article

Caching Efficiency Enhancement at Wireless Edges with Concerns on User’s Quality of Experience

Feng Li, Kwok-Yan Lam, Li Wang, Zhenyu Na, Xin Liu, and Qing Pan

1 School of Computer Science and Engineering, Nanyang Technological University, 639798, Singapore
2 School of Electronic Science and Engineering, Nanjing University, Nanjing 210093, China
3 College of Information Engineering, Zhejiang University of Technology, Hangzhou 310023, China
4 School of Information Science and Technology, Dalian Maritime University, Dalian 116026, China
5 School of Information and Communication Engineering, Dalian University of Technology, Dalian 116024, China

Correspondence should be addressed to Li Wang; liwang2002@zjut.edu.cn

Received 24 November 2017; Revised 29 December 2017; Accepted 31 December 2017; Published 30 January 2018

Copyright © 2018 Feng Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Content caching is a promising approach to enhancing bandwidth utilization and minimizing delivery delay for new-generation Internet applications. The design of content caching is based on the principles that popular contents are cached at appropriate network edges in order to reduce transmission delay and avoid backhaul bottleneck. In this paper, we propose a cooperative caching replacement and efficiency optimization scheme for IP-based wireless networks. Wireless edges are designed to establish a one-hop scope of caching information table for caching replacement in cases when there is not enough cache resource available within its own space. During the course, after receiving the caching request, every caching node should determine the weight of the required contents and provide a response according to the availability of its own caching space. Furthermore, to increase the caching efficiency from a practical perspective, we introduce the concept of quality of user experience (QoE) and try to properly allocate the cache resource of the whole networks to better satisfy user demands. Different caching allocation strategies are devised to be adopted to enhance user QoE in various circumstances. Numerical results are further provided to justify the performance improvement of our proposal from various aspects.

1. Introduction

In today’s rapid development of Internet applications, users have ever-increasing demand for communication bandwidth and yet with high expectation for delivery speed from the Internet. In this connection, scholars proposed content-centric networking (CCN) to fit in the explosive development of Internet services [1–4]. With CCN, the network infrastructure implements a unified naming scheme for content data such that the content names are mapped to the IP in the middle layer. This way, CCN makes the content the main unit of network addressing and communication. To achieve the high bandwidth utilization and minimize delivery delay, the operation of CCN is based on a built-in caching function within the network wherein intermediate network nodes cache all passing content and hence potentially can respond to user requests without obtaining the requested content from their sources. Among others, the caching strategy has a direct impact on the quality of CCN content distribution network performance and hence attracted growing attention of the network research community in recent years.

Technically, with CCN, many of the basic principles for traditional Internet design are no longer applicable to the current network environment. Future Internet should have functions or services to support transparent separation between content name from the name of the host on which the content resides. Besides, network security and quality of service should also be part of the built-in attributes of the future network. Research on future Internet is pursued in accordance with two main development trends: (1) the evolutionary route, which is based on the smooth evolution of the existing IP network architecture by introducing network services incrementally in a transparent and backward-compatible manner, and (2) the revolutionary idea, which...
aims to revamp the Internet with a new network architecture with the necessary next-generation infrastructure services. In the long run, incremental improvements will introduce a lot of complexity and inefficiency to the network infrastructure and hence will not be able to fundamentally solve the problems of flexibility, reliability, and controllability of the Internet architecture. In this paper, we focus on the evolutionary route by investigating an IP-based caching network for meeting needs of new-generation Internet applications [5–9].

The caching technique of Leave Copy Everywhere (LCE), as one kind of path caching scheme, is one of the most popular caching strategy [10, 11], whose basic concept is to cache the content data by all the caching nodes on return paths when receiving the content data. While the caching nodes do not have enough space, the replacement policy of least recently used will be performed to update local saved content which means the longest unused contents should be replaced with the new-coming contents. Although the above-mentioned strategy is simple and easy to deployed, it can likely cause redundancy of caching data and high frequency of caching replacements. Thus, how to design an appropriate caching strategy, reducing caching redundancy and increasing cache hit, is a critical point which is worthy of being focused on [12, 13].

To solve the caching issue raised above, many researchers have proposed different improvement methods, such as the LCD strategy proposed in [14]. When the data is responded, it is only cached in the next hop node of the hit node. Reference [15] mentioned the ProbCache strategy, in which nodes along the way with a probability \( P \) cache content, and the closer to the requester, the greater the value of \( P \). In addition, Cho et al. in [16] proposed WAVE Strategy; the upstream node obtains the file popularity by recording the number of times that the file is requested. Furthermore, according to the value of the popularity to the downstream nodes, the proposal suggested the need to cache chunk and its number. However, the above-mentioned strategies mainly focus on the reduction of caching redundancy and the increase of cache hit ratio and less consider the network bandwidth savings. Literature [17] proposed Betw caching solution. In the proposed method, when the packet is returned along the requested path, the caching is only made at the most important node along the way. This strategy reduces the number of hops in the content transmission in the network. However, the excessive concentration of caching contents makes the high-center nodes overloaded which will inevitably lead to frequent replacement of caching content, thereby reducing the subsequent utilization of caching resource. In [18], the authors proposed and described the main building blocks of a hybrid Fog-supported P2P architecture for wireless content delivery which is named the Fog-Caching P2P architecture. The proposal explored the topological information locally available at the serving Fog-caching nodes, with the intention to speed up the data searching process carried out by the served pairs. In [19], the authors raised an alternative architecture in which the caching and routing operations are separated and the caching routers are dedicated to packet routing when the caching server controls the whole network caching contents. During the course, the caching server is supposed to be installed at every domain.

Many emerging techniques and mathematical methods, including convex optimization, game theory, and intelligent algorithms, are utilized to solve the optimal caching issues in wireless networks [20–23]. Furthermore, numbers of solutions and strategies from other areas in communication networks can inspire the scholars to solve the caching difficulties [24–30]. In this paper, we focus on the IP-based caching architecture which is a research hot point in recent years [31–35]. In [31], the authors proposed an in-network caching solution in IP-based networks by inserting a content identifier into a new-coming IPv6 extension header, wherein the proposed architecture is called caching-based IP networks. In [32], the authors studied the information-centric networking architectures with caching functions for their potentials to adapt to emergency communications which is involved in the satellite communication networks. In [33], a new type of threat was introduced to the coherence-exploiting hardware Trojan which can be implanted in master components in a system by taking advantage of cache coherence protocols. In [34], the authors proposed mandatory content access control, a distributed information flow control mechanism enables a content provider to control which network nodes can cache its contents.

In this paper, we propose a caching strategy to enhance caching efficiency in IP-based network architecture by introducing the notion of quality of user experience (QoE) and simplify the replacement method of neighbor node’s caching. We first design a simple cooperative caching strategy in IP-based networks to build and discover suitable neighbor caching nodes and determine whether caching the new content data according to the weight information of the contents and current caching status. The caching replacement method is performed within one-hop node region to facilitate the operations. Furthermore, we mainly pay our attention on the enhancement of caching efficiency by considering the effects of caching resource placement on the user QoE. Due to the nonlinear relationship between the increase of caching resource and the user QoE, detailed research on the change of user QoE deserves more attention. In this case, we give the definition of QoE with respect to the caching allocation. We mainly focus on how to efficiently use the caching resource in condition of no enough caching space for the requests of wireless edges. Various solutions to respond to the situations where different user demands and emergency degree are raised to address the caching resource allocation. In additions, we provide the simulated results to testify our proposals.

The main contributions of this paper are as follows:

1. A cooperative caching strategy is designed to simplify the establishment and implementation of neighbor caching and replacement in case of limited caching resource being required by wireless nodes. The main contribution of this paper lies in that we propose a simplified caching solution with high feasibility and low complexity.
(2) A solution for caching efficiency enhancement is devised by introducing wireless node QoE. We investigate how to use limited caching space to increase the user QoE which leads to an enhancement of caching efficiency as a result.

The rest of this paper is organized as follows. We give the system model of caching at wireless edges in Section 2. Then, the scheme for caching replacement on neighbor nodes along with the strategy to enhance caching efficiency is proposed in Section 3. In Section 4, numerical results are supplied to evaluate the performance of our proposed algorithm. At last, we conclude this paper in Section 5.

2. System Model

In this paper, we consider an IP-based information-centric caching networks where caching is installed at plenty of wireless edges as shown in Figure 1. Due to the problems on compatibility and routing performance of information-centric-based network mechanism, totally discarding traditional IP-based networks and adopting information-centric networks are uneasy to fulfill and deploy currently. On the other hand, for the multimedia-oriented network circumstance, the communication mode of device-to-device transmission in IP-based networks has the defects of unavailability for wireless caching. Thus, we consider an IP-based information-centric networks with many edges caching as a hybrid networks to smooth the compatibility and enhance network performance as well. In Figure 1, some of the wireless edges are equipped with caching modules rather than all of them. In fact, it is not essential for all the terminals or base stations to be cached so as to enhance network performances. Only some key nodes with large relay transmission tasks should be installed by caching modules.

Besides, as information-centric caching networks are characterized by their large numbers of content resource and dynamic network mechanism, it is difficult to deploy them as centralized systems. There also exists feasibility problem for caching networks to completely replace traditional IP-based networks. Therefore, we consider a hybrid approach combining traditional networks and caching networks to perform our proposed caching strategy for enhancing caching efficiency and improving system performance. The basic diagram of transmitting signal and caching content in the proposed model is as shown in Figure 2.

In the scenario described in Figure 2, users can decide whether they should carry IP option according to demands in application services. When the data packets carrying IP information arrive at usual routers, they can be relayed directly without any logical processing. When the data packets reach the caching routers, they will be processed by logical units to judge in the routers whether part of contents should
3. Enhancement of Caching Efficiency

Wide deployment of caching at wireless edges can reduce redundancy flow and content transmit-delay. However, due to very limited space of caching especially at personal terminals, how to efficiently utilize caching space of kinds of edges to improve caching efficiency and network performance is a critical issue.

Currently, classical caching strategy of Leave Copy Everywhere (LCE) may lead to large caching redundancy and frequent caching replacement, which inspires more practical cooperative caching techniques to be raised. Since the path cooperative caching solution tends to perform decision-making based on every single path, wherein the caching policy can incur collision when numbers of paths will pass by same node. Meanwhile, the centralized processing mode in local cooperative caching scheme proposed recently will cause relatively large communication overhead.

To solve the problems mentioned above, this paper proposes a simple reactive cooperative caching method with the consideration of user’s quality of experience to enhance caching efficiency when no enough caching space can be used during the course. Our proposal carries out caching and replacement only in one hop of adjacent neighbor router. The cooperative caching method should fulfill the foundation and discovery of neighbor relation and then make a judgement on whether caching the new-coming content data according to local and neighbor caching information in the past. A cooperative replacement will occur when local caching space is insufficient and the information exchange with neighbor router has been completed.

3.1. Cooperative Caching. The IP-based cooperative caching strategy we propose is a reactive real-time policy for caching content data. The strategy can contain the following processes as follows:

1. Find neighbor and foundation between caching routers.
2. When the caching routers in this circumstance receive new content data, they will check neighbor caching information tables. If the caching tables have the corresponding contents, no caching will be performed; otherwise, the caching operation will be forwarded according to storage space. If the storage space is sufficient enough, the router caches the content. Otherwise, replacement strategy should be implemented.

3.1.1. Mechanism of Neighbor Finding and Foundation. In general, the routers within one-hop physical distance can be called neighbor nodes. For the proposed IP-based caching networks, usual routers and caching routers coexist simultaneously as shown in Figure 3. For traditional mode of broadcasting to find neighbor routers, information overflow and high communication cost are likely to happen. Thus, we design a simple and efficient neighbor finding mechanism in this paper. In this section, we propose an IP address-based replacement scheme which adopts Request Integration Table (RIT) to build the neighbor relationship between caching routers to ensure symmetry of both uplink and downlink data and find neighbor routers. The diagram is given in Figure 4.

The main feature of the IP address-based discovery and replacement mechanism for caching router proposed in this paper is that each caching node accomplishes this by continuously updating and recording neighbor nodes’ status information within one-hop scope. In the caching discovery process, the current cache node first determines whether the currently received data has a caching value and checks whether its own caching status is suitable for storing the content. If all the above conditions are satisfied, the current node can complete the content caching and continue forwarding the related content to the next node. If the current node discovers that the data content does not need to be cached, it is forwarded directly to the next node.
With the massive data services, the caching spaces of each variety of users of data information is flooding. Compared does not meet the need of real-time transmission of caching tent. That is to say, the overall caching space of the network need to make a certain choice of the caching content. This case, how to cache efficiently is very important.

The traditional methods to improve the efficiency of caching content include dividing the content of the cache or dividing the priority, the emergency degree, and the reasonable cleaning of the historical cache information. This paper intends to raise a method to improve caching efficiency from the perspective of user satisfaction. According to current network load, caching allocation strategy for traditional network to the user group assumes that the cell users’ QoS demand is stable and similar and allocates from several fixed cache sizes and neighbor nodes to users. In actual mobile network interaction, the behavior characteristics and habits of different users are different. The communication QoS psychological expectation on the time axis and the change of user satisfaction will affect the final results of the caching scheduling. That is to say, there is a certain deviation between mobile network’s payment on caching resource and the user satisfaction of expected revenue. In this case, how to further improve the satisfaction index according to the slight changes of customer satisfaction, which is to further improve the efficiency of cache use, in the case of limited resources, is worthy of in-depth study.

Firstly, assuming that the completion time of the mobile user expects the task to be completed, the total time length of the task from submission to feedback confirmation is \( T_c \). If the task is completed within time \( T_c \), the task can obtain high satisfaction. The minimum caching resource needed to complete this task is the unit execution resource, marked as \( C_i \). The central controller processes \( n \) tasks successively in a time segment \( T_r \). At the same time, we introduce the quality of user experience to represent the user’s experience satisfaction, and the quality of the total user experience after the execution of each task can be expressed as

\[
Q = \sum_{i=1}^{n} \text{QoE}_i,
\]

where QoE represents the quality of user experience gained after the execution of \( i \) task. User QoE is closely related to the expected completion time of users. When \( T_c - T_e \leq 0 \) (\( T_c \) is the actual completion time of tasks), the QoE value is the highest, otherwise getting down gradually. However, in general, in the process of calculation of task execution, the relationship between the two factors is not simply linearly decreasing but a ladder-like broken line, as shown in Figure 5. It can be seen from the ladder curve that the system completes the task time reduction and sometimes cannot directly improve the user QoE, or increasing the amplitude is not proportional to growth. Then, according to the nonlinear characteristics of the user QoE, more reasonable scheduling of caching resources can increase the utilization of network resources.

For given overall caching capacity \( C \) in the whole networks, then the contribution degree to QoE with per caching resource can be given as

\[
U_{\text{profit}} = \frac{\sum_{i=1}^{n} \text{QoE}_i}{C}.
\]
Initialization:
Caching contents \( V = \{ v_1, v_2, \ldots, v_n \} \) and corresponding weight factor \( W = \{ w_1, w_2, \ldots, w_n \} \);
Weight threshold \( W^t \); Number of caching contents \( k \);
Weight comparison factor \( g_i \);
New caching content's size \( s \);
Neighbor router available cache size \( a \);

Main routine:
Calculate comparison factor \( g_i = \frac{w_i}{w_j}, i \) and \( j \in n \);
Reorder the comparison factor of new caching contents as \( G = \{ g'_1, g'_2, \ldots, g'_n \} \)
and the caching content's size set as \( S = \{ s'_1, s'_2, \ldots, s'_n \} \);

if \( g \leq g'_i \) (\( i = 1, 2, \ldots, n \)), then
The new content data will not be cached by the local caching node;

if \( w_i \leq W^t \), then
The new content data will not be cached by the local caching node;
else
if \( s \leq s'_i \) \& \( a! = 0 \), then
Replace the content correlated with \( g'_i \) by the new content;
else
\( S_u = s'_i \);
\( F_g = 1 \);
for \( j = k - 1 \) to 1 run
if \( g > g'_j \), then
\( S_u = S_u + s'_j \);
if \( s \leq S_u \) \& \( a! = 0 \), then
Replace queue of \( G = \{ g'_1, g'_2, \ldots, g'_n \} \) by the new content data;
\( F_g = 0 \);
Break;
end if
else
break;
end if
end for
if \( F_g == 1 \), then
The new content will not be cached by the caching node;
end if
end
end

Algorithm 1: Caching replacement algorithm.

Figure 4: IP-based router replacement mechanism.
In order to maximize the contribution degree in fixed time slot $T_s$, the following condition should be met

$$QoE = \max \left( \sum_{i=1}^{n} F \left( T_{e_i} - B_i \right) \right),$$  

(3)

where

$$C = B_1 + B_2 + \ldots + B_n.$$  

(4)

Through gathering and analyzing the characteristics information of user behaviors, we can obtain the distribution probability function of load task and then further predict sets $C_i$ and $T_{e_i}$ for time slot $T_s$. Thus, convert the optimization objective mentioned above to be a multidimensional and linear optimization problem to solve.

From the aspect of systems, the allocation strategy of caching resource mainly depends on the two factors including the distribution probability of user task delivery over time line and completion time expected by the user. We plan to divide the service time in task controller into numbers of equal time slots $T_1, T_2, \ldots, T_m$. The average execution time for single task can be expressed as

$$dt = \frac{\left( \sum_{k=1}^{s} \text{time (task)}_k \right)}{k}.$$  

(5)

The user’s expected completion time can be judged by the user’s feedback results or the frequency of the task to be refreshed. If the frequency of user query or refresh is faster, the shorter the expected task completion time is, the more urgent the demand is, the higher the satisfaction expectation is. According to the user’s query frequency, the expected degree of the time slot $T_{e_i}$ can be divided into three priority levels and marked by different colors. Blue means lower than the average, white means average, and red means that the frequency of inquiries is greater than the average.

The strategy of caching resource supply should be carefully divided according to the degree of urgency and the busy state. When the user’s work status is classified into two types, idle and busy, the query frequency corresponds to the general and emergency two cases. The caching scheduling policy at this point is shown in Figure 6. Different optimization strategies should be applied to solve the problem of caching resource allocation at the wireless edges. In the following section of simulation test, we will give the performances of our proposal compared with the other general method of caching solution.

4. Numerical Results

In this section, we perform several simulation tests to testify our proposed method. Based on the NS-3 simulation platform [36]; we simulate and analyze the proposed cooperative caching replacement solution within neighbor nodes and the scheme of caching efficiency improvement. In the following tests, a network platform is built based on numbers of caching module, and a BRITE topology generator is used to generate a network topology according to Waxman probability model [37, 38]. The network includes overall 80 routers, containing 20 caching routers. The location of all the wireless edges is random. Furthermore, 60 terminals along with 10 content servers coexist in the network. The user requests to send signal by the rate of 10 times per second. The contents of the network data have 5000 types. Interaction between neighbors is performed per 3 seconds; the total simulation time is 2000 seconds.

As shown in Figure 7, we give the performance of cache hit ratio with changing cache size between our proposal and other two methods. LCE solution denotes the caching technique of Leave Copy Everywhere which is a traditional method to caching the content data by all the caching nodes along the return paths when receiving the content data. The method of CCTH denotes the cooperative caching strategy with neighbor detection scope of two hops. CCTH adopts similar caching scheme compared with our proposal except for the detection scope. We can observe from Figure 7 that the cache hit ratio rises obviously with increase of cache size. Due to more caching space occupied in network nodes, the performance of LCE outperforms our proposed method. However, the increase of cache hit ratio of our proposal is relatively faster than that of LCE. Besides, the cache hit ratio
of CCTH method is approaching to that of LCE solution and better than our method due to expanded detection scope from one neighbor hop to two neighbor hops.

Furthermore, as shown in Figure 8, the traffic reduction ratio with changing cache size is presented. It is apparent from the figure that the traffic capacity will be affected due to the increase of cache size. Since the LCE method needs more caching nodes to involve in the networks, the traffic reduction ratio is relatively high compared to other methods. Besides, due to the increase of scanning area of neighbor caching replacement for CCTH method, more replacements and exchanges of caching contents will occur which leads to the increase of traffic reduction ratio as a result.

In Figure 9, we further give the performances of user QoE under optimal cache allocation raised in this paper and average caching allocation. The definition of QoE has been given in (3), and we randomly select 10 terminals to figure out the average value. We thus allocate and activate more cache resources to the terminals raising frequent requests during the transmission. The optimal caching allocation elastically assigns limited resource to users. Compared to the average cache allocation method, our scheme can receive high user QoE with less cache size. Besides, as shown in Figure 10, the cache hit ratio with changing cache size is presented. With the increase of cache size, the cache hit ratio in optimal cache allocation scheme is approaching to the method of LCE who has more caching nodes in the networks.
5. Conclusions

In this paper, we investigate the caching problems including caching replacement and caching efficiency enhancement in IP-based communication systems where most of the caches are installed in various wireless edges. The main contribution of this paper lies in that we devised a simplified caching replacement and user QoE-based caching allocation scheme to improve caching efficiency at wireless edges. During caching replacement operation, each caching node accomplishes the cache detection by continuously updating and recording neighbor nodes’ status information within one-hop scope. Besides, we introduce the concept of user’s QoE to describe the caching efficiency from another aspect. By adopting different kinds of optimization solutions to properly allocate nodes’ cache resources, user’s QoE can be improved as a result. Simulation tests are performed to testify our proposal’s performances. By comparing with other traditional caching method in wireless systems, we give the presentations of several solutions in perspective of cache hit ratio and traffic reduction ratio. With the increase of cache size, our proposal can reap rapid and reasonable capacity improvement with relatively low cache size.

Disclosure

Li Wang is currently at School of Computer Science and Engineering, Nanyang Technological University, Singapore 639798.

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work was partly supported by the National Natural Science Foundation of China under Grants 51404211 and 61601221 and partly supported by BAE Systems.

References


