

Research Article

An Optimal Routing Algorithm in Service Customized 5G Networks

Haipeng Yao,¹ Chao Fang,¹ Yiru Guo,² and Chenglin Zhao³

¹State Key Lab of Networking and Switching Technology, Beijing University of Posts and Telecommunications, P.O. Box 199, No. 10, Xitucheng Road, Haidian District, Beijing 100876, China

²Beijing University of Posts and Telecommunications, P.O. Box 199, No. 10, Xitucheng Road, Haidian District, Beijing 100876, China

³Key Lab of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, P.O. Box 199, No. 10, Xitucheng Road, Haidian District, Beijing 100876, China

Correspondence should be addressed to Haipeng Yao; yaohaipeng@bupt.edu.cn

Received 5 November 2015; Accepted 21 December 2015

Academic Editor: Xin Wang

Copyright © 2016 Haipeng Yao 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.

With the widespread use of Internet, the scale of mobile data traffic grows explosively, which makes 5G networks in cellular networks become a growing concern. Recently, the ideas related to future network, for example, Software Defined Networking (SDN), Content-Centric Networking (CCN), and Big Data, have drawn more and more attention. In this paper, we propose a service-customized 5G network architecture by introducing the ideas of separation between control plane and data plane, in-network caching, and Big Data processing and analysis to resolve the problems traditional cellular radio networks face. Moreover, we design an optimal routing algorithm for this architecture, which can minimize average response hops in the network. Simulation results reveal that, by introducing the cache, the network performance can be obviously improved in different network conditions compared to the scenario without a cache. In addition, we explore the change of cache hit rate and average response hops under different cache replacement policies, cache sizes, content popularity, and network topologies, respectively.

1. Introduction

Mobile communication is the fastest growing field in the telecommunications industry, where the cellular radio network is the most successful mobile communication system. A new mobile generation has appeared approximately every 10 years since the first 1G system was introduced in 1982. The first 2G system was commercially deployed in 1992, and the first 3G system appeared in 2001. 4G systems fully compliant with IMT Advanced were first standardized in 2012. With the widespread use of Internet, mobile hosts overtake fixed ones, not only in terms of numbers but also in terms of traffic load [1]. How to practically deal with the explosive growth in wireless traffic and meet the increasing mobile users' needs becomes a growing concern in the current cellular networks due to the increasing network cost [2, 3]. Recently, 5G networks have been designed by fully considering the ideas related to future network, for example, Software Defined Networking (SDN) [4], Content-Centric Networking (CCN)

[5, 6], and Big Data to provide simply faster speeds and meet the needs of new use cases, such as the Internet of Things as well as broadcast-like services and lifeline communication in times of natural disaster.

SDN is an emerging network architecture where network control is decoupled from forwarding. This migration of control is formerly tightly bound in individual network devices and now enables the underlying infrastructure to be abstracted for applications and network services, which can treat the network as a logical or virtual entity. The main idea of SDN is to allow software developers to depend on network resources in the same easy way as they do on storage and computing resources. In SDN, the network intelligence is logically centralized in software-based controllers, and network devices become simple packet forwarding devices that can be programmed via an open interface [4]. Based on SDN, the system can provide a global view of network and control the forwarding traffic according to the mobile users' demands.

CCN is a receiver-driven, data-centric communication protocol [7, 8]. All communication in CCN is performed using two distinct types of packets: *Interest packet* and *Data packet*. Both types of packets carry a name, which uniquely identifies a piece of data that can be carried in one Data packet. Besides, to receive data the users requested, each CCN content router maintains three major data structures: a *Content Store (CS)* for temporary caching of received Data packets, a *Pending Interest Table (PIT)* to contain the name of the Interest packet and a set of interfaces from which the matching Interests have been received, and a *Forwarding Information Base (FIB)* to forward the Interest. CCN has recently emerged as one of the most promising architectures for the diffusion of contents over the Internet. A major feature of this novel networking paradigm is in-network caching [9, 10], which can cache content objects to shorten the distance of user requests. When the content is sent in reply to user requests, it can be cached by any CCN content router along the way back to the request originators. With in-network caching, CCN can provide low dissemination latency and reduce network load, as requests no longer need to travel until the content source but are typically served by a closer CCN content router along the routing path [11].

To cope with the explosive increase of mobile data and make a timely response to the users' requests and network problems, more and more researchers have paid more attention to Big Data. Big Data is a set of techniques and technologies that require new forms of integration to uncover large hidden values from large datasets that are diverse, complex, and of a massive scale, which makes possible the centralized network control, and timely processing and analysis of massive traffic in wireless networks.

In order to resolve the problems traditional cellular radio networks face, some advantages related to future network have been considered. In this paper, we propose a service-customized 5G network architecture by introducing the ideas of separation between control plane and data plane, in-network caching, and Big Data processing and analysis and design an optimal routing algorithm for this architecture. The main contributions of this paper are as follows.

- (i) We propose a novel service-customized 5G network architecture, which fully considers the benefits brought by separation between control plane and data plane, in-network caching, and Big Data processing and analysis.
- (ii) We design an optimal routing algorithm and abstract it as a optimal model in the proposed 5G network architecture, which can meet a mobile user's request with minimal network latency, and realize the load balance in the network.
- (iii) Simulation results reveal that, by introducing the cache, the network performance can be obviously improved in different network conditions compared to the scenario without a cache. In addition, we explore the change of cache hit rate and average response hops under different cache replacement policies, cache sizes, content popularity, and network topologies, respectively. In the simulation, we use

three main popular Least Frequency Used (LFU), Least Recently Used Policy (LRU), and Random (RND) cache replacement policies [12, 13] to evaluate the performance of the system.

The rest of this paper is organized as follows. In Section 2, the novel service-customized 5G network architecture is presented. In Section 3, the optimal energy-efficient routing model in the proposed architecture is given. Simulation results are presented and discussed in Section 4. Finally, we conclude this study in Section 5.

2. Service-Customized 5G Network Architecture

As shown in Figure 1, the service-customized 5G network architecture introduces the in-network cache in the network devices, such as base stations and routers, realizes the separation between control plane and data plane, and adds the Big Data processing and analysis functions to the control plane.

From Figure 1, we can see that the introduced cache in the network devices can buffer the contents that mobile users are interested in and place the contents near the users. The following same content request can be satisfied by the cache without transmitting to the source server. Moreover, the separation between control plane and data plane makes the system have a global view of resources (e.g., network, compute, and cache) and dynamically configure the underlying network equipment in time by using the online and off-line Big Data processing and analysis platform.

The workflow of the proposed architecture is as follows. Initially, the controller keeps monitoring the network and updating information in a fixed period. Therefore, it can obtain the number of users' requests, the real-time loading situations. When a mobile consumer requests a content, the request with the content's name is encapsulated and forwarded to the network edge device. Then the controller uses the collected information to calculate the optimal routing path with the minimum network cost among the providers which cache the content the user requests. After that, it updates the in-network cache status based on the number of users' requests and cache replacement policies. Based on the advantages of Big Data platform, the controller can timely obtain the optimal routing path and update the cache status.

To satisfy QoS of a VoD application, the application firstly tells the controller what kinds of and how many resources (e.g., network bandwidth, storage capacity) it needs in a request packet, which may be constructed in some way. After receiving the request packet, the controller uses the Big Data processing platform to analyze the resource information contained in the packet and then automatically allocate resources according to the demands of the application. Finally, a virtual network is formed and the proposed optimal routing algorithm model is used to achieve minimal network latency under constrained conditions by monitoring the dynamic storage and content status.

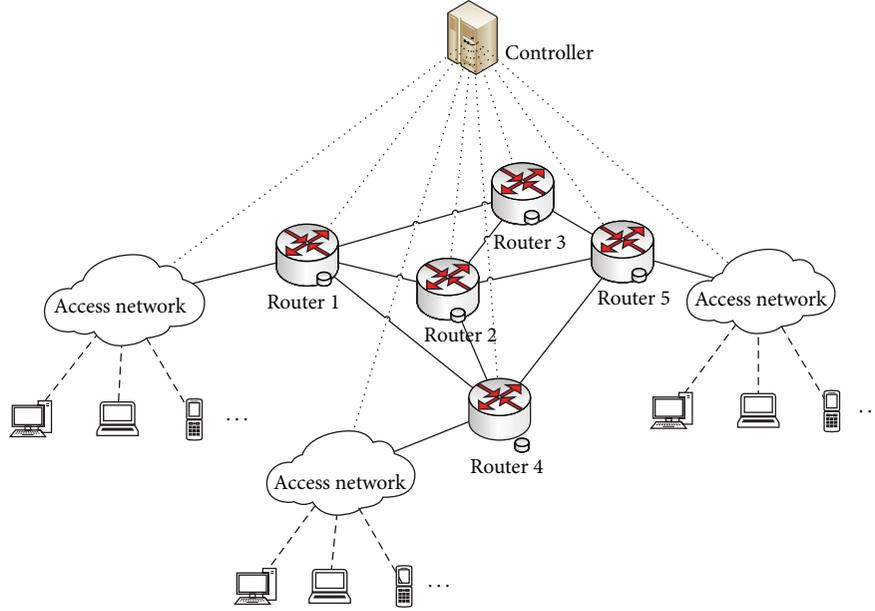


FIGURE 1: Service-customized 5G network architecture.

3. Optimal Routing Algorithm Model

We model the network as a connected graph $G = (V, E)$, where $V = (V_1, V_2, \dots, V_N)$ is the set of content routers in the network, and $E \subseteq V \times V$ is the set of network bidirectional links. Let $O = (O_1, O_2, \dots, O_M)$ be the set of content objects that can be available in the network. All of the objects are initially distributed in the network servers, which are directly connected to edge content routers [14]. For the sake of readability, the term “content router” and “node” will be used interchangeably here.

In this paper, our objective is to achieve minimal network latency by addressing the question of how each content router with limited caching capacity caches contents in the network. Therefore, the optimal routing problem can be formulated as an integer linear problem (ILP) as follows:

$$\begin{aligned}
 \min_{x,y}: & \sum_{i=1}^N \sum_{k=1}^M q_i^k \sum_{j=1}^N d_{ijk} y_{ijk} \\
 \text{s.t.}: & \sum_{j=1}^N y_{ijk} = 1, \quad \forall i, k \\
 & y_{ijk} \leq x_{jk}, \quad \forall i, j, k \\
 & \sum_{k=1}^M s^k x_{ik} \leq C_i, \quad \forall i \\
 & y_{ijk} \in \{0, 1\}, \quad x_{ik} \in \{0, 1\}, \quad \forall i, j, k,
 \end{aligned} \tag{1}$$

where q_i^k is the request rate for object O_k at node i , d_{ijk} is the distance consumed by node i to request content object O_k from node j , C_i is the maximum cache size at router i , and s^k is the size of the content object O_k . Moreover, x_{ik} takes

the value of 1 if node i caches a copy of element k , and 0 otherwise. y_{ijk} takes the value of 1 if node i downloads a copy of content object k from node j , and 0 otherwise. Obviously, some methods (e.g., genetic algorithm) can be used to obtain the optimal solution.

4. Simulation Results and Discussions

In this section, we use computer simulations to evaluate the performance of the structure of the new architecture. We first describe the simulation settings and then present the simulation results and compare them.

4.1. Simulation Settings

4.1.1. Network Topologies. The simulation is carried out in the power-law and transit-stub topology, respectively. Power-law topology is generated by Inet topology generator [15] and includes 64 content routers, where 40 edge routers are connected to the users and makes the users' number $T = 40$. Transit-stub topology is generated by using GT-ITM library [16] and has 24 routers, where 15 edge routers are connected to the users and sets the users' number $T = 15$.

4.1.2. Input Data. In the simulation, there are 100 different contents, and the total number of content objects is 2×10^6 in the network. We assume each object has the same size, and the content popularity follows the Zipf distribution and the wide range of the skewness factor α is 0.5–1.5 [11, 17, 18].

4.1.3. Cache Size. In the simulation, we abstract the cache size for each CCN content router as a proportion that the cache size is defined as the relative size to the total amount of different contents in the network. Given that the cache size

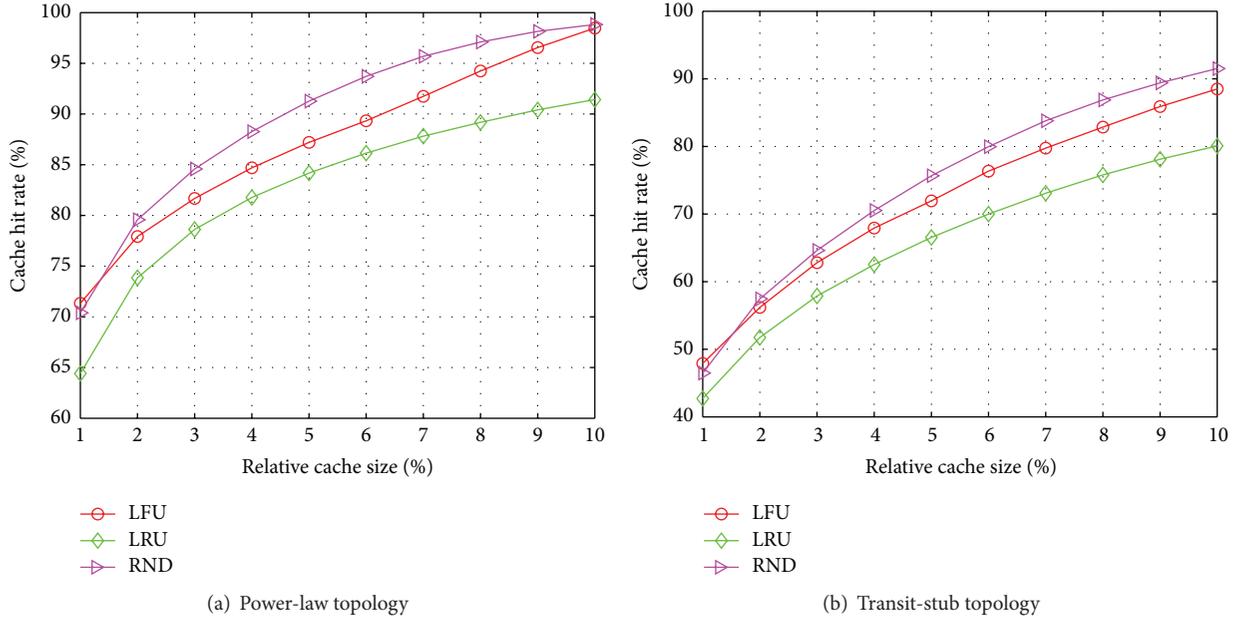


FIGURE 2: Cache hit rate versus cache size under different network topologies.

of the CCN router is small in realistic network [13, 19–23], we evaluate the network performances for each caching scheme when the cache memory size varies from 1% to 10% [23].

4.1.4. Comparative Policy. The widely used caching replacement policies in CCN is Least Frequency Used (LFU), Least Recently Used Policy (LRU), and Random (RND) cache replacement policies. Therefore, we use three main popular LFU, LRU, and RND cache replacement policies to evaluate the performance of the proposed architecture and designed routing policy in the simulation.

4.1.5. Performance Metrics. In the simulation, we evaluate cache hit rate (CHR) and average response hops (ARH), which are the two important metrics to measure network QoS.

- (i) CHR is the ratio of the amount of requested objects served by routers rather than the source servers in the system to the total amount of requested contents.
- (ii) ARH is referred to as the average number of the routers traversed by the response packets from the source servers or routers to the requesting mobile users.

4.2. Performance Evaluation Results. Figure 2 shows the cache hit rate of each solution with varying cache sizes under different network topologies when content popularity is 0.8. From Figure 2, we can see that the cache hit rate of each policy changes in the same way as cache size increases in the power-law and transit-stub topology, respectively. However, the cache hit rate under power-law topology is better than that under transit-stub topology because the total cache size is much larger. Moreover, the performance of RND is best

while that of LRU is worst. The reason is that RND policy fits the global network users' behavior well by random replacing the contents in the cache. But for LRU, it replaces the content in the cache with the recently accessed one, which increases the contention of each cache and further reduces the cache hit rate of the system. LFU can achieve a high cache hit rate by caching the frequently contents while its performance is worse than RND due to slowly catching up with the change of content popularity of each node.

Figure 3 shows the average response hops of each solution with varying cache sizes under different network topologies when content popularity is 0.8. From Figure 3, we can see that the proposed architecture can obviously reduce average response hops by introducing in-network cache. Moreover, average response hops of each policy changes in the same way as cache size increases in the power-law and transit-stub topology, respectively. However, the average response hops under power-law topology is worse than that under transit-stub topology because of the larger distance of the network nodes. Moreover, the performance of LFU is best while that of LRU is worst. The reason is that each node using LFU policy caches the contents with high access frequency, which makes the contents that users are interested in available near them. But for LRU, it leads to the frequent content replacement in each node, which reduces the average response hops of the system. Although LFU can achieve the best cache hit rate, it cannot make the interested contents cached near the users, which obtains a worse average response hops compared with LFU policy.

Figure 4 shows the cache hit rate of each solution with varying content popularity under different network topologies when cache size is 5%. From Figure 4, we can see that the cache hit rate of each policy changes in the same way as content popularity increases in the power-law and

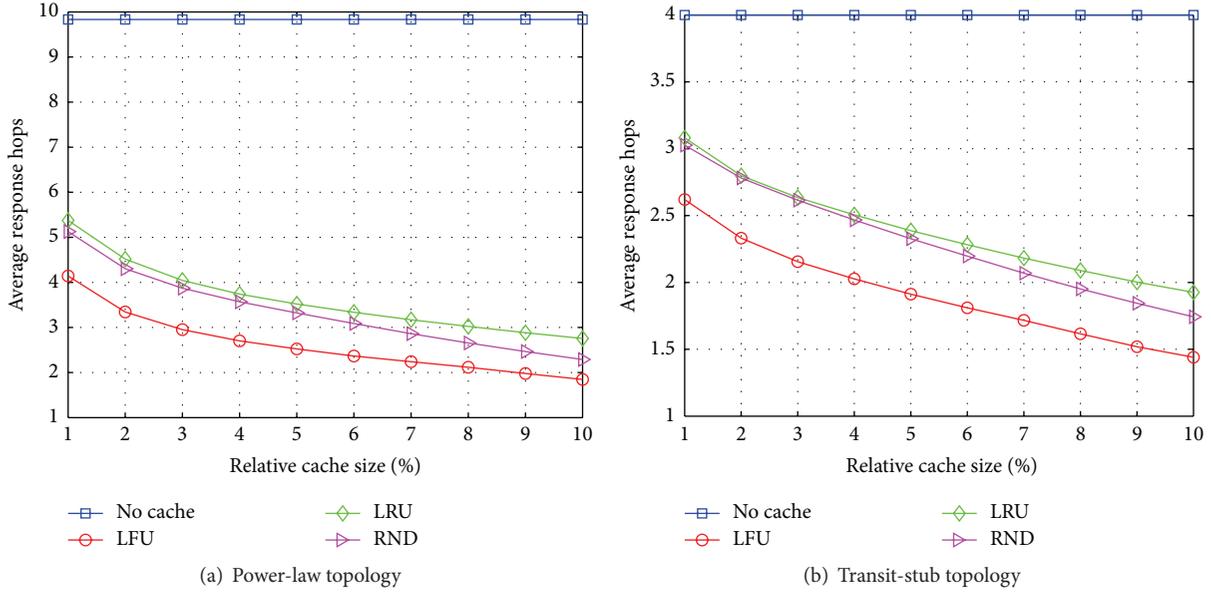


FIGURE 3: Average response hops versus cache size under different network topologies.

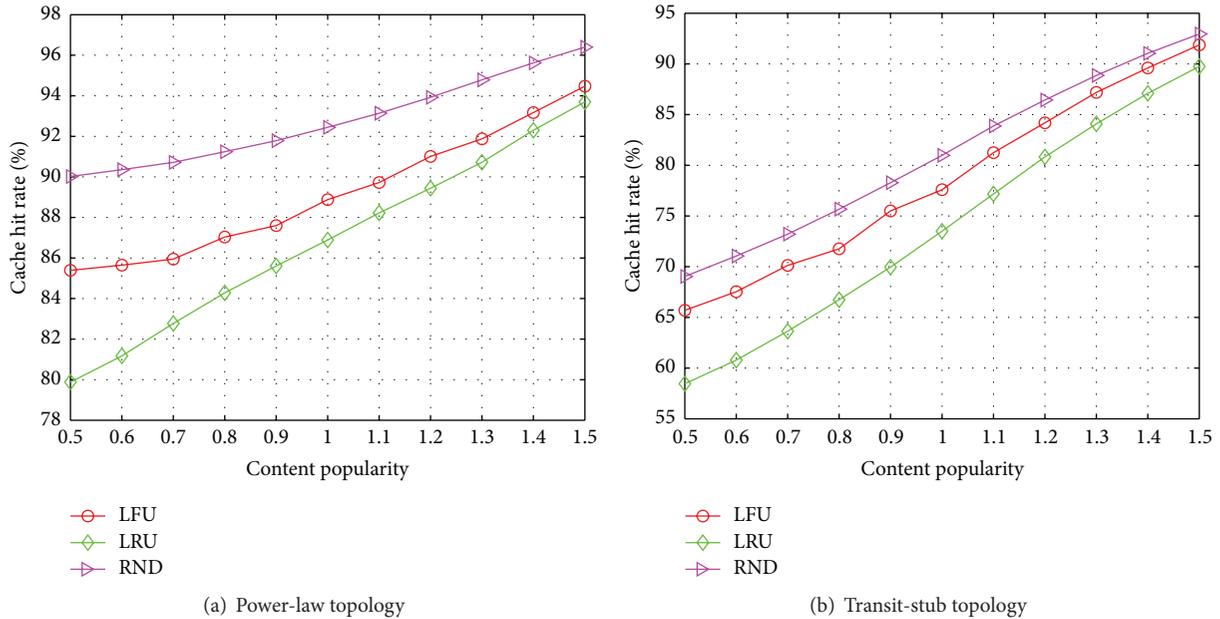


FIGURE 4: Cache hit rate versus content popularity under different network topologies.

transit-stub topology, respectively. The reason is that the increasing content popularity gradually reduces the number of content types in the network. However, the cache hit rate under power-law topology is better than that under transit-stub topology because the total cache size is much larger. Moreover, the performance of RND is best while that of LRU is worst. The reason is similar to that of Figure 2, because the fact that the number of content types is reduced means the cache size increases relatively.

Figure 5 shows the average response hops of each solution with varying content popularity under different network

topologies when cache size is 5%. From Figure 5, we can see that the proposed architecture can obviously reduce average response hops by introducing in-network cache. Moreover, average response hops of each policy changes in the same way as content popularity increases in the power-law and transit-stub topology, respectively. However, the cache hit rate under power-law topology is worse than that under transit-stub topology because of the larger distance of the network nodes. Moreover, the performance of LFU is best while that of LRU is worst. The reason is similar to that of Figure 3, which has been mentioned in the previous paragraph.

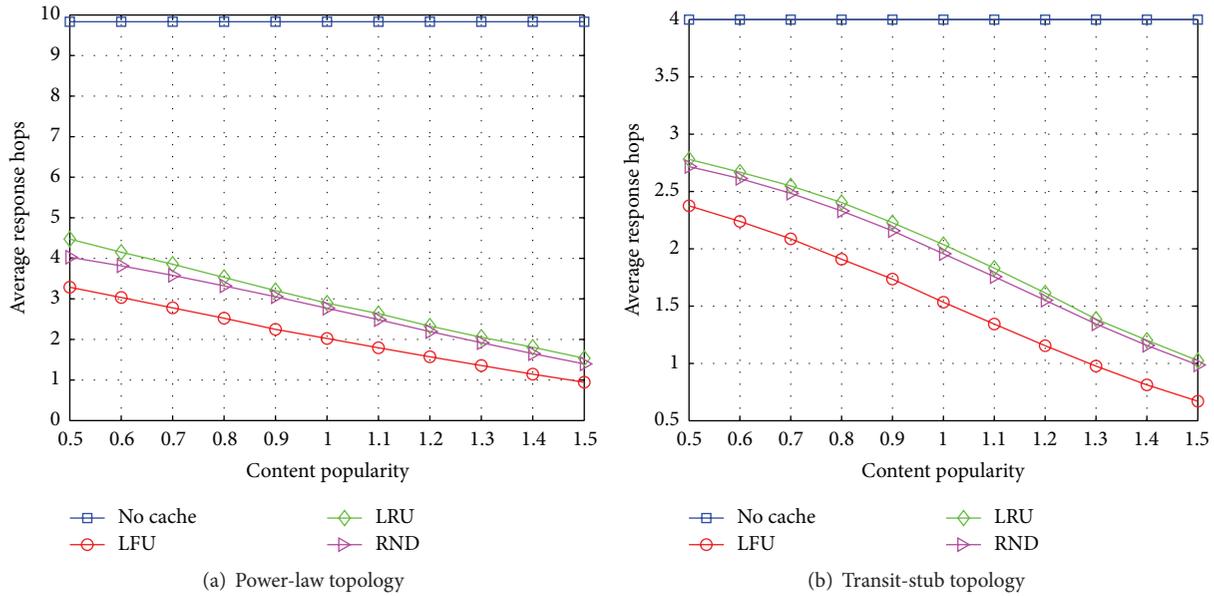


FIGURE 5: Average response hops versus content popularity under different network topologies.

5. Conclusions and Future Work

In this paper, by adopting the advantages of separation between control plane and data plane, in-network caching, and Big Data processing and analysis, we propose a service-customized 5G network architecture to overcome the problems current cellular radio networks face. Moreover, we design an optimal routing algorithm for this architecture, which can minimize average response hops in the network. Simulation results reveal that, by introducing the cache, the network performance can be obviously improved in different network conditions compared to the scenario without a cache. In addition, we explore the change of cache hit rate and average response hops under different cache replacement policies, cache sizes, content popularity, and network topologies, respectively.

With recent advances of wireless mobile communication technologies and devices, more and more end users access the Internet via mobile devices, such as smart phones and tablets. Therefore, we will study user mobility in the proposed model in the future. Moreover, it would be interesting to discuss how to find an online routing algorithm to minimize network cost in the future work.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

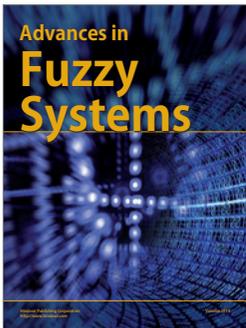
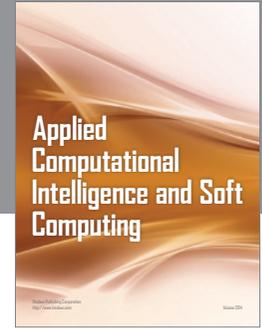
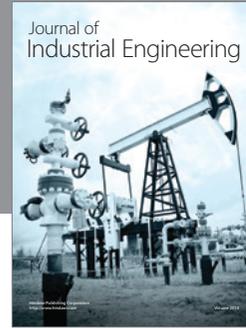
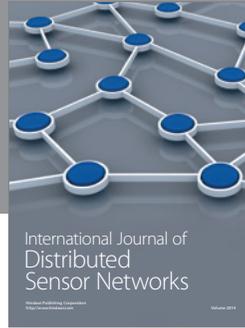
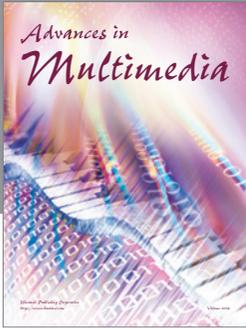
Acknowledgment

This work was supported by NSFC (61471056) and China Jiangsu Future Internet Research Fund (BY2013095-3-1, BY2013095-3-03).

References

- [1] G. Xylomenos, X. Vasilakos, C. Tsilopoulos, V. A. Siris, and G. C. Polyzos, "Caching and mobility support in a publish-subscribe internet architecture," *IEEE Communications Magazine*, vol. 50, no. 7, pp. 52–58, 2012.
- [2] J. Araujo, F. Giroire, Y. Liu, R. Modrzejewski, and J. Moulhierac, "Energy efficient content distribution," in *Proceedings of the IEEE International Conference on Communications (ICC '13)*, pp. 4233–4238, Budapest, Hungary, June 2013.
- [3] V. Mathew, R. K. Sitaraman, and P. Shenoy, "Energy-aware load balancing in content delivery networks," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 954–962, IEEE, Orlando, Fla, USA, March 2012.
- [4] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turetli, "A survey of software-defined networking: past, present, and future of programmable networks," *IEEE Communications Surveys & Tutorials*, vol. 16, no. 3, pp. 1617–1634, 2014.
- [5] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of green information-centric networking: research issues and challenges," *IEEE Communications Surveys & Tutorials*, vol. 17, no. 3, pp. 1455–1472, 2015.
- [6] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of energy-efficient caching in information-centric networking," *IEEE Communications Magazine*, vol. 52, no. 11, pp. 122–129, 2014.
- [7] C. Yi, A. Afanasyev, I. Moiseenko, L. Wang, B. Zhang, and L. Zhang, "A case for stateful forwarding plane," Tech. Rep., University of Arizona, University of California, Los Angeles, University of Memphis, 2012.
- [8] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A distributed energy consumption optimization algorithm for content-centric networks via dual decomposition," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '14)*, pp. 1848–1853, Austin, Tex, USA, December 2014.

- [9] I. Psaras, R. G. Clegg, R. Landa, W. K. Chai, and G. Pavlou, "Modelling and evaluation of CCN-caching trees," in *NET-WORKING 2011: 10th International IFIP TC 6 Networking Conference, Valencia, Spain, May 9–13, 2011, Proceedings, Part I*, vol. 6640 of *Lecture Notes in Computer Science*, pp. 78–91, Springer, Berlin, Germany, 2011.
- [10] G. Carofiglio, V. Gehlen, and D. Perino, "Experimental evaluation of memory management in content-centric networking," in *Proceedings of the IEEE International Conference on Communications (ICC '11)*, Kyoto, Japan, June 2011.
- [11] N. Choi, K. Guan, D. C. Kilper, and G. Atkinson, "In-network caching effect on optimal energy consumption in content-centric networking," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 2889–2894, Ottawa, Canada, June 2012.
- [12] V. Jacobson, D. K. Smetters, J. D. Thornton, M. F. Plass, N. H. Briggs, and R. L. Braynard, "Networking named content," in *Proceedings of the ACM Conference on Emerging Networking Experiments and Technologies (CoNEXT '09)*, pp. 1–12, Rome, Italy, December 2009.
- [13] D. Rossi and G. Rossini, "Caching performance of content centric networks under multi-path routing (and more)," Tech. Rep., Telecom ParisTech, Paris, France, 2011.
- [14] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "An energy-efficient distributed in-network caching scheme for green content-centric networks," *Computer Networks*, vol. 78, pp. 119–129, 2015.
- [15] H. Chang, R. Govindan, S. Jamin, S. J. Shenker, and W. Willinger, "Towards capturing representative as-level internet topologies," *Computer Networks*, vol. 44, no. 6, pp. 737–755, 2004.
- [16] E. W. Zegura, K. L. Calvert, and M. J. Donahoo, "A quantitative comparison of graph-based models for internet topology," *IEEE/ACM Transactions on Networking*, vol. 5, no. 6, pp. 770–783, 1997.
- [17] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: evidence and implications," in *Proceedings of the 18th IEEE Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, vol. 1, pp. 126–134, IEEE, New York, NY, USA, March 1999.
- [18] K. Katsaros, G. Xylomenos, and G. C. Polyzos, "MultiCache: an incrementally deployable overlay architecture for information-centric networking," in *Proceedings of the IEEE Conference on Computer Communications Workshops (INFOCOM '10)*, pp. 1–5, IEEE, San Diego, Calif, USA, March 2010.
- [19] Z. Li, G. Simon, and A. Gravey, "Caching policies for in-network caching," in *Proceedings of the 20th International Conference on Computer Communications and Networks (ICCCN '11)*, Maui, Hawaii, USA, July-August 2011.
- [20] H. Xie, G. Shi, and P. Wang, "TECC: towards collaborative in-network caching guided by traffic engineering," in *Proceedings of the IEEE Conference on Computer Communications (INFOCOM '12)*, pp. 2546–2550, Orlando, Fla, USA, March 2012.
- [21] J. Li, H. Wu, B. Liu et al., "Popularity-driven coordinated caching in named data networking," in *Proceedings of the 8th ACM/IEEE Symposium on Architectures for Networking and Communications Systems (ANCS '12)*, pp. 15–26, Marina del Rey, Calif, USA, October 2012.
- [22] J. Li, B. Liu, and H. Wu, "Energy-efficient in-network caching for content-centric networking," *IEEE Communications Letters*, vol. 17, no. 4, pp. 797–800, 2013.
- [23] N. Laoutaris, S. Syntila, and I. Stavrakakis, "Meta algorithms for hierarchical web caches," in *Proceedings of the 23rd IEEE International Performance, Computing, and Communications Conference (IPCCC '04)*, pp. 445–452, IEEE, Phoenix, Ariz, USA, April 2004.



Hindawi

Submit your manuscripts at
<http://www.hindawi.com>

