IEEE Access
Multidisciplinary : Rapid Review : Open Access Journal

Received May 15, 2017, accepted June 6, 2017, date of publication June 19, 2017, date of current version September 27, 2017.

Digital Object Identifier 10.1109/ACCESS.2017.2715824

The Collaboration for Content Delivery and Network Infrastructures: A Survey

QINGMIN JIA[1], RENCHAO XIE[1,2], TAO HUANG[1,2], JIANG LIU[1,2], AND YUNJIE LIU[1,2]
[1]State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China
[2]Beijing Advanced Innovation Center for Future Internet Technology, Beijing 100124, China

Corresponding author: Tao Huang (htao@bupt.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61501042, in part by the National High Technology Research and Development Program (863) of China under Grant 2015AA016101, and in part by the Beijing Nova Program under Grant Z151100000315078.

ABSTRACT With the explosive growth of Internet traffic and the rapid development of network technology, content delivery has become a significant service in the current Internet. However, the existing content delivery solutions, such as peer-to-peer and content delivery network, have many insufficient aspects. Meanwhile, the collaboration between content delivery and network infrastructures has been considered as a promising technique in network field. From the perspective of collaboration, content delivery systems can make full use of the network characteristics and the effective information provided by the network operators, so as to improve the efficiency of the content distribution and optimize the overall performance of the network. In this paper, we present a comprehensive survey on the collaboration for content delivery and network infrastructures. First, we provide some of the works, which have been done on collaboration solutions from two perspectives: evolutionary and revolutionary. And then, the advantages and disadvantages of these solutions are compared and analyzed from three aspects of technology, business, and standardization. Finally, we outline some challenges and research directions in the future.

INDEX TERMS Collaboration, content delivery, network architecture, network optimization.

I. INTRODUCTION

With the continuous development of many emerging technologies, such as mobile Internet, cloud computing, big data, Internet of Things and so on, the demand for Internet traffic is increasing rapidly. According to the Cisco Visual Networking Index (VNI) report 2016 [1], global IP traffic has increased more than fivefold in the past 5 years, and will increase nearly threefold over the next 5 years. Meanwhile, consumer Internet video traffic will be 82 percent of all consumer Internet traffic in 2020, up from 70 percent in 2015. In addition, content delivery network traffic will deliver nearly two-thirds of all Internet video traffic by 2020.

The existing network faces great challenges due to the rapid growth of Internet traffic, especially video traffic. On the one hand, the network becomes extremely congested, and the response for user request becomes quite slow. In order to improve the network conditions, Internet service providers (ISPs) have to constantly expand the network bandwidth. However, with the reduction of voice and short messaging service (SMS) business as well as the drop of traffic price, ISP network bandwidth expansion does not bring considerable benefits for themselves. On the other hand, the content providers are also faced with enormous pressure. For example, every 400ms delay in search responses will result in a 0.59% drop in users' search requests for Google [2], and every 100ms increase in latency will cut profits by 1% for Amazon [3], [4].

To better cope with the current rapid growth of network traffic and the challenges brought by the network congestion, improving the efficiency of content delivery has been considered as a significant approach. And the research of content delivery has also attracted a lot of attention in the academia and industry, such as peer-to-peer network, content distribution network, etc. In the following, we briefly introduce these two kinds of content delivery technologies.

Peer-to-peer (P2P) is a distributed content delivery technology based on overlay network [5], [6]. Compared with the current Client-Server (C/S) model, each peer in the P2P network has a dual role for both the server and the client, which can provide services to other peers and get services from other peers. The P2P system intrinsically supports content distribution, because when a peer publishes a content request, all the peers owned objective content will respond. Thus, the content delivery is extremely efficient in P2P networks. Consequently, P2P technique has been widely used in the Internet, especially in file sharing.

Content delivery network (CDN) is also an important distributed network [7], [8], which is composed of a large number of edge servers (also called surrogates) distributed in different regions. Moreover, the CDN can offer fast and reliable services and applications for users. More specifically, according to specific content delivery rules or policies, such as cache decision, cache update, server load balancing and so on, the CDN can push/pull the contents (including Internet sites, online video, online games and other content sources) to the edge servers located close to users, thus the users can obtain the desired content nearby. Therefore, CDN is an important technique to ease the Internet network congestion, improve the response speed, and optimize the user experience.

Although the P2P and CDN have achieved great success in technology and industry, they still have a lot of challenges. On the one hand, the P2P applications occupy a large amount of network bandwidth resources, and bring great pressure to network operators. At the same time, the quality of service for non-P2P network applications is difficult to guarantee. On the other hand, CDN has less underlying network information which can optimize the distribution performance, such as the ISP network topology, link load information and user location, resulting in a high cost and low efficiency. Therefore, it is necessary to explore efficient content distribution schemes.

In order to further improve the efficiency of content delivery, the collaboration for content delivery and network infrastructures has been proposed [9], [10]. Through the perception of network topology and link load information, it can optimize the transmission path of content distribution and the selection of content servers. Thus it can reduce the time delay for request and response, and improve user quality of experience (QoE).

In addition to the traditional content delivery solutions, recently, the clean-slate approaches such as software-defined networking (SDN) [11]–[13], information-centric networking (ICN) [14], [15] have emerged, which inherently have advantages in content delivery. At the same time, the emergence of the new architectures brings new opportunities and challenges, resulting in the promotion of collaboration possibility for content delivery and network infrastructures. Hence, it is necessary to explore the collaborative revolutionary network architectures, aiming to deploy as early as possible and improve the performance of content delivery.

Recently, many excellent survey papers on content delivery technologies and network infrastructures have been done. In [16], Passarella focuses on the traditional content-centric technologies, and provides a survey for the traditional content delivery solutions including P2P and CDN. In [17], Lu *et al.* focus on CDN-P2P-hybrid architecture technology, and provide an analysis and comparison for CDN-P2P hybrid content delivery systems and models. In [18], Frank *et al.* also mainly focus on traditional content delivery technologies, and provide an overview of collaboration for content delivery and network infrastructures. Furthermore, the revolutionary network technologies also attract a lot of attention. In [19], Zhang *et al.* study the caching techniques in information centric networking, and present a survey of caching techniques from the perspective of ICN performance optimization. And Zhang *et al.* [20] also present a comprehensive survey of caching techniques for ICN, which pay more attention to the recently proposed caching mechanisms for ICN. In addition, Fang *et al.* [21] provide a survey on ICN caching techniques in terms of energy-efficiency.

Although these works on content delivery technologies and network infrastructures are extremely excellent, most of them do not comprehensively describe the collaboration solutions and fully analyze the advantages and disadvantages. In this survey, we focus on the collaboration for content delivery and network infrastructures, and aim at providing an overview of the recent developments in the collaboration solutions and discussing about research issues and future directions. Moreover, we divide the collaboration solutions into evolutionary solutions and revolutionary solutions: the former collaborates with the traditional content delivery solutions and the latter collaborates with the emerging content delivery solutions.

A structure of the paper is given in Fig. 1. The rest of this article is organized as follows: we present the evolutionary collaboration solutions in Section II, and then the revolutionary collaboration solutions are described in Section III. Next, we compare and analyze the collaboration solutions in Section IV. In Section V, the challenges and future development directions are discussed. Finally, a brief summary and our conclusion are provided in Section VI.

II. EVOLUTIONARY COLLABORATION FOR CONTENT DELIVERY AND NETWORK INFRASTRUCTURES

In this section, we mainly present the evolutionary collaboration solutions, which adopt traditional content distribution technologies, such as P2P, CDN etc. In addition, we also consider transparent caching and collaborative caching, which play the important roles in the collaboration for content delivery and network infrastructures. These solutions are summarized in Table 1.

A. THE COLLABORATION SOLUTION FOR P2P AND ISP

Content distribution system based on P2P technology can greatly reduce the distribution cost of content providers and improve the scalability of the system. However, the mismatch between the overlay network and the underlying network leads to the waste of network resources, resulting in a large amount of redundant traffic and serious network congestion, which intensifies the contradiction between P2P content providers and ISPs. Thus, many ISPs take a lot of measures to limit the traffic generated by P2P applications, which influences the QoE of end-users. Therefore, how to effectively utilize the bandwidth resources of the underlying network and

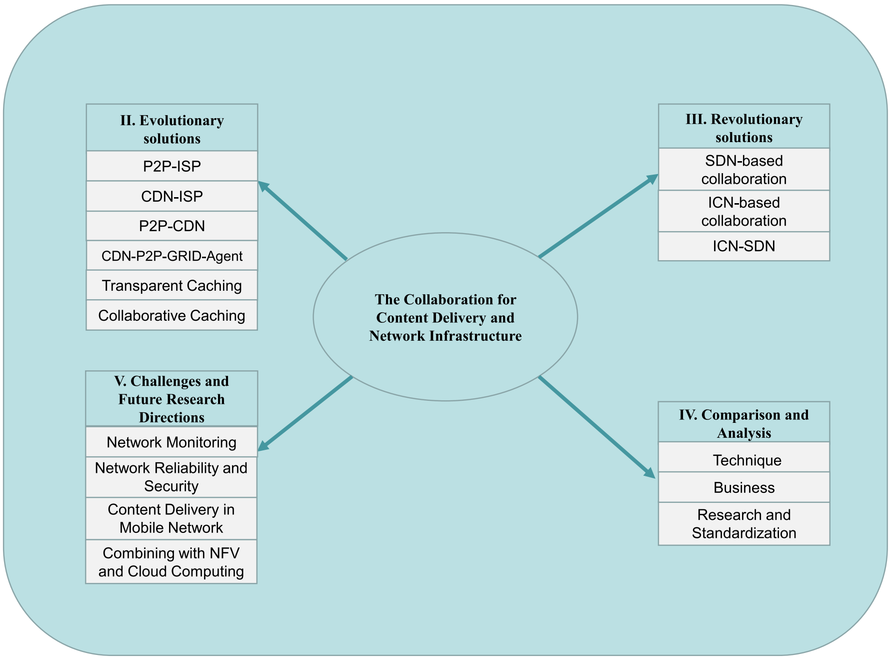


FIGURE 1. Roadmap of the collaboration for content delivery and network infrastructures.

to reduce the traffic pressure of the P2P content distribution system to the ISP network is the key of the P2P technology to be healthy and sustainable development. Recently, this issue has aroused considerable attention of academia and industry, resulting in promotion of collaboration research for P2P and ISP.

P2P and ISP collaboration, that is, the ISP provides the service of network topology and state information, and P2P content provider optimizes its topology and content scheduling by accessing the service. Under this framework, ISP can achieve a variety of services by controlling the granularity of network topology and state information, as well as privacy protection by anonymous processing. Meanwhile, the P2P content provider can reduce the cost of topology detection, and avoid the risk of being blocked by ISP. In this subsection, we review the existing collaboration solutions for P2P and ISP.

In [9], Aggarwal *et al.* propose the P2P Oracle network architecture, which can enable ISP to help P2P to optimize the selection of service nodes, so as to improve the performance of P2P. Because the P2P system relies on the application layer routing based on overlay network topology, the routing of P2P system and the topology of the underlying Internet are separated. Therefore, it is very difficult to manage and optimize the P2P traffic using the traditional traffic engineering. Moreover, the huge P2P traffic brings great pressure to the operators, and has a negative influence on non-P2P application. Accordingly, in the Oracle scheme, the P2P user supplies a list of potential P2P neighbors to the oracle system operated by ISP, then the Oracle system ranks them according to certain metrics. Thus, the P2P node can choose a more reasonable adjacent nodes based on the rank list. Hence, P2P traffic can get more reasonable guidance by the Oracle system. At the same time, the operators can use this mechanism to better manage the huge P2P traffic, such as maintaining the P2P traffic localization, optimizing the P2P traffic flow path, thereby improving the network performance.

Moreover, Xie *et al.* propose another P2P and ISP collaboration scheme named P4P [22], which provides extremely effective cooperative traffic control between P2P applications and network providers. In the traditional P2P network, the selection of content nodes and transmission of content are random, which leads to inefficient network resource usage and low application performance. Accordingly, the P4P provides the iTrackers, which consist of multiple interfaces for networks to communicate with P2P applications. Moreover, the iTrackers operated by network providers can provide a lot of information regarding the network status. And the P2P clients can query the iTracker to obtain a number of desired information regarding the network providers, such as policy, distance and network capability. According to the information, P2P clients choose the better neighbor nodes and transmission links. Therefore, the P4P can effectively utilize the network bandwidth resources, reduce the backbone network transmission pressure and operating costs, and improve the performance of P2P applications.

The principle of P2P technology is based on the exchange of popular content, and the more popular the content is, the higher the distribution efficiency of P2P becomes. However, for the unpopular content, the distribution effect of P2P is not outstanding, and it is easy to consume a large amount of bandwidth resources. Even using the P4P technology,

TABLE 1. Overview of evolutionary collaboration solutions.

Collaboration	Solutions	References	Contributions
P2P-ISP	Oracle	Aggarwal *et al.* [9]	Optimizing the P2P traffic by ranking the potential P2P neighbors, and thus helping the P2P networks make better choices in picking neighboring nodes.
	P4P	Xie *et al.* [22]	Providing information regarding the network status for P2P applications, and guiding the P2P client to choose the better neighbor nodes and transmission links.
CDN-ISP	PaDIS	Poese *et al.* [10] [23]	Providing rank lists regarding end-user assignment to servers based on current network conditions and accurate network locations, and guiding the end-user to choose the most preferable CDN servers.
	CaTE	Poese *et al.* [24]	The main idea of CaTE is the CDN gives a list of potential servers for a users request and the ISP returns a rank of these servers based on the network condition,so as to optimize both content delivery performance and link utilization.
	NetPaaS	Frank *et al.* [25]	NetPaaS is the evolution of CDN-ISP collaboration solution, which adapts the development of technology and industry, especially the Network Function Virtualization (NFV) [26].
P2P-CDN	PeerCDN	J. Wu *et al.* [27]	Proposing a layered architecture called PeerCDN, which combines P2P and CDN seamlessly with their inherited excellent feature.
	HCDN	H. Jiang *et al.* [28]	Hybrid content distribution network (HCDN) architecture is similar to PeerCDN.
	Ono	Choffnes *et al.* [29]	Proposing a novel approach to reduce the costly cross-ISP traffic without sacrificing system performance.
CDN-P2P-GRID-Agent	UPGRADE-CDN and UPGRADE-CN	G. Fortino *et al.* [30] [31] [32]	Proposing a content delivery framework integrating CDN, P2P, GRID and Agent technologies.
Transparent caching	In-line	PeerApp UltraBand [33]	The characteristic of In-line system is deployed in the transmission link.
	Out-of-band	Qwilt [34]	Out-of-band system is deployed outside the network traffic path, which separates the control plane and data plane.
Collaborative caching	collaboration model or solution in wired networks	G. Dán *et al.* [35], J. Dai *et al.* [36], G. Fortino *et al.* [37] [38]	Proposing cooperative caching models or content delivery scheme
	collaboration model or solution in wireless networks	F. Pantisano *et al.* [39], Z. Chen *et al.* [40], Khreishah *et al.* [41], J. Dai *et al.* [42], Y. Xu *et al.* [43]	Proposing collaborative caching schemes or models in wireless network
	collaboration solution for CDN and mobile network	N. Loulloudes *et al.* [44], F. Z. Yousaf *et al.* [45], D. Munaretto *et al.* [46]	Proposing cooperative solutions for CDN and mobile network

the distribution of unpopular content is still not efficient. Therefore, the collaboration between P2P and ISP mainly aims at popular content distribution.

B. THE COLLABORATION SOLUTION FOR CDN AND ISP

Nowadays, a large amount of Internet traffic is carried by CDN, and CDN has become the one of the most significant content delivery technologies. However, the CDN still exists a lot of issues, which influence the efficiency of content delivery. One of the issues for CDN is the lack of the awareness for the network conditions. Namely, the CDN has to dynamically map end-users to appropriate servers without being fully aware of the network conditions within an ISP or the end-user location. As a result, the CDN and ISP are in such a situation. On the one hand, ISP has the network topology and link load status information, but the content distribution capability is insufficient, and the network is very congested. On the other hand, CDN is difficult to achieve the optimal content routing and distribution. Therefore, the collaboration between ISP and CDN is an extremely important approach to improve the performance of content delivery.

The collaboration for CDN and ISP can achieve triple-win results. For CDN, it can obtain the network information that can help the CDN improve performance, such as network topology, link load and user location. So it is not required to carry out large-scale network measurement and topology detection. For ISP, it can reduce the traffic pressure and gain better traffic management and network utilization, thereby reducing the cost of investment and operation. For users, they can get a better network experience. In the following section, we give an overview of the collaboration for CDN and ISP.

To cope with the challenge driven by CDN, Poese *et al.* propose a Provider-aided Distance Information Systems (PaDIS) in [10] and [23]. The method is that PaDIS lets an ISP influence server selection of CDN by extending its DNS infrastructure. In more detail, PaDIS can discover content servers' location diversity, and monitor information regarding network state including topology information and connectivity information, thus it can maintain an up-to-date annotated map of the ISP network. Accordingly, PaDIS can rank lists of available servers based on the server diversity and up-to-date network state. According to the server lists, PaDIS can recommend CDN select optimal servers, thus the CDN can assign the best servers to users for end-user performance.

In addition to PaDIS, Poese *et al.* also propose a Content-aware Traffic Engineering (CaTE) scheme in [24], in essence, a collaboration scheme between ISP and CDN. The approach is that the CDN gives a list of potential servers for a users request and the ISP returns a rank of these servers to optimize both content delivery performance and link utilization. The architecture of CaTE system is similar to PaDIS system, they both have network monitoring component and query processing component. However, the big difference between them is whether the CDN content servers is open. Namely, in CaTE system, CDN provides a list of the potential content servers they operate but in PaDIS system CDN doesn't. Note that the collaboration for CDN and ISP is only in server selection, not routing.

In [25], Frank *et al.* propose a CDN-ISP collaboration solution called NetPaaS (Network Platform as a Service), a system aiming at providing accurate user-server assignments and in-network server allocations for the CDN. In this architecture, NetPaaS mainly consists of network monitoring component, informed user assignment component, and server allocation interface component. Moreover, NetPaaS can be considered the evolution from the CaTE, because a lot of NetPaaS extensions including the components and functions are based on CaTE. In particular, NetPaaS combines with virtualization technology [47], [48] and the ISP can offer virtual machines (VMs) [49] to CDN. Therefore, NetPaaS allows the CDN to expand or shrink its footprint inside the ISP network on demand. In summary, NetPaaS allows CDN and ISP to cooperate not only on user assignment, but also on dynamically deploying and removing servers and thus scaling content delivery infrastructure on demand.

CDN technology has inherent advantages on content delivery, and through the network topology and link load information provided by ISP, it can improve the content distribution efficiency, greatly reduce the congestion of backbone network. But CDN technology is still based on the traditional IP architecture and routing mechanism as well as forwarding strategy. Thus, network controllability and flexibility is poor. Moreover, CDN is not free from the C/S structure, and CDN server is easy to become the bottleneck of the system.

C. THE COLLABORATION SOLUTION FOR CDN AND P2P

CDN and P2P are two important content distribution technologies, which have their own advantages and disadvantages. CDN can reduce the delay of the content delivery to the user, but it needs to deploy a large number of network edge cache servers, and thus the cost is higher. P2P can avoid the deployment cost, but P2P applications usually occupy a large amount of network bandwidth resources, and bring great pressure to network operators. In addition, it can reduce the quality of service due to the departure of the P2P nodes. Accordingly, the collaboration of CDN and P2P can achieve the complementary advantages and improve the overall performance of content distribution. In the following section, we introduce several solutions of the collaboration for CDN and P2P.

In [27], Wu *et al.* propose a novel hybrid architecture called PeerCDN, which combines P2P and CDN with their inherited excellent features. PeerCDN is a two-layer streaming architecture. Upper layer is a server layer which is composed of original CDN servers including origin servers and cache servers. And the lower layer consists of groups of clients who request the streaming services, and each client is considered as a client peer in the group. Each group of client peers is led by the nearby replica server. And these client peers contribute their resources through the coordination of the leader peer.

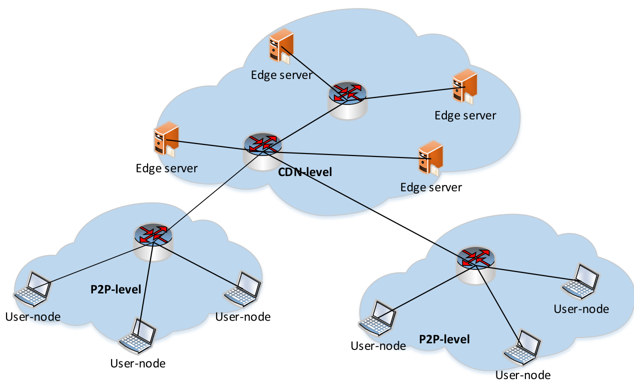


FIGURE 2. The architecture of HCDN.

Moreover, Jiang *et al.* propose a hybrid content distribution network (HCDN) in [28] and [50], a hierarchical network architecture which is similar to PeerCDN, can achieve the large-scale content distribution. As shown in Fig. 2, the idea of the HCDN architecture is that the CDN system is set up in the core network, which is responsible for distributing the contents of the source server to the network edge proxy server. The P2P system is set up in the edge network, which can be used to obtain the contents from the proxy server or P2P node. The scheme makes full use of the advantages of the two technologies, and makes up the shortcomings of each single technology. In summary, this scheme reduces the load of CDN proxy server and reduce the cost of deployment, and it can also improve the quality of network service.

In [29], Choffnes *et al.* propose a novel approach to reduce the costly cross-ISP traffic without sacrificing system performance. And this approach can recycle network views gathered at low cost from CDNs to drive biased neighbor selection without any path monitoring or probing. Moreover, the implementation named Ono is designed as a plugin for compatibility with the Azureus BitTorrent client.

In addition to aforementioned solutions, there are also a lot of relevant research achievements [51]–[55]. Despite the approaches are different, the objectives are consistent: make full use of the advantages of CDN and P2P, improve the performance of content delivery, reduce the cost and ease network congestion.

Through the collaboration between P2P and CDN, the content distribution of backbone network utilizes CDN technology, and the content distribution of local network utilizes P2P technology. Hence, this collaboration scheme has the advantages of both P2P and CDN, and it also integrates the disadvantages of both.

D. THE COLLABORATION SOLUTION FOR CDN, P2P, GRID AND AGENT

In addition to the collaboration solution for CDN and P2P, joint consideration for the CDN, P2P, GRID and Agent is another significant collaboration method [30], [56]. In order to make the distribution, management, discovery and delivery of content more efficient and robust in CDN, it is necessary to enhance the collaboration for CDN, P2P, GRID and Agent. Because the P2P technology can increase the dynamism and fault-tolerance, the GRID [57] can favour robustness and multi-organization application, and the Agent [58] can foster intelligent behaviour and self-organization features.

In [30], Fortino and Russo propose the UPGRADE-CDN framework, which uses P2P, GRID and Agent for the development of CDN. In the proposed solution, UPGRADE-CDN consists of three core components. The first component is CoDelivery, which is a P2P based cooperative system for the content delivery to the clients. The second component is GRedirector, which is a GRID and DNS based redirection system to redirect the client requests to the most suitable surrogate. The third component is AMonitor, which is an Agent-based system for monitoring the status of surrogates, the access network, and the distribution network. In addition, Fortino and Mastroianni [31] further extend the concept of CDN, and propose the Content Networks (CNs). CDN mainly focus on the infrastructure and mechanisms for content deliver. However, the Content Networks can not only support the content delivery, but also support the content creation, modification, placement and management. Hence, the Content Networks can better meet the new requirements for network applications and services. Moreover, the P2P, GRID and Agent can also be applied in the Content Networks (i.e., UPGRADE-CN) to improve the performance [32].

With the development of CDN technology, the development and deployment of new services and components in CDN becomes complex and costly. It is necessary to prototype, monitor, and predict the behavior of new CDN services/components in a controlled simulated environment. Hence, in [59], Fortino *et al.* propose an agent-based approach to design and analyze the CDN. In addition, the agent-based extensible CDN framework can provide high-level programming abstractions and tools, which can support the simulation of different CDN architectures and mechanisms and allow for automatic evaluation of three main CDN performance indices: average user perceived latency, cache hit ratio and utility.

Integrating P2P, GRID and Agent technologies in CDN framework is a significant method to improve the performance of the content delivery system. Through the collaboration of these technologies, the content delivery system can become more efficient, robust and intelligent.

E. THE TRANSPARENT CACHING

Transparent caching (TC) is a novel distributed network managed by ISP, which also reflects the collaboration for content delivery and network infrastructures. Similar to the traditional web caching, the transparent caching also has a large number of cache servers deployed close to end-users, which identify, store and deliver the most popular Internet content [60]. However, the transparent caching maintains end-to-end client-server relationship compared to the web cache. In detail, end-user initial request can reach content origin servers through transparent caching, and the server can get the information regarding the end-user, such as IP address, terminal type, browser and so on. The origin server response contents can also reach the end-user after inspection by transparent caching, and the end-user can get the related information of the content origin server. Hence, the end-users and content providers can not perceive the existence of caching server, which is the reason why it is called transparent caching. In essence, transparent caching is collaboration solution for cache and network. Therefore, we consider the transparent caching as a kind of collaboration solutions for content delivery and network infrastructures. According to the deployment position of cache system, transparent caching can be classified into two types: in-line and out-of-band. In the following section, we introduce the two types of transparent caching solutions.

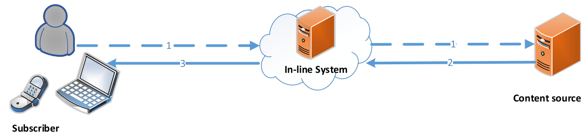


FIGURE 3. **the In-line architecture.**

In-line system is a significant transparent caching solution deployed in the transmission link close to the end-users. And all the data traffic that is traversing the network passes through the inline device, allowing the device to inspect the content and operate. PeerApp UltraBand is a kind of in-line transparent caching solution [33]. It consists of a cache engine which is used to detect repeatedly requested content, storage disks which hold the content, and a management center which is used for reporting, configuration and management. As shown in Fig. 3, end-user requests the object content from the origin content server and established session, and the PeerApp UltraBand system inspects the request and passes the request to origin content server. Then the origin content server executes content delivery logic (authorization, content adaptation, reporting, etc.) and delivers the requested object content. Finally, PeerApp UltraBand inspects response content header and payload. If object content is already in Ultraband cache server, the object content will be served directly from cache server, and origin content server will be directed to stop serving the object. Otherwise, object content will be delivered from the origin content server and stored for future use.

Out-of-band system is another significant transparent caching solution which is deployed outside the network traffic path. It can obtain a copy of the network traffic through data replication mechanisms such as optical taps or

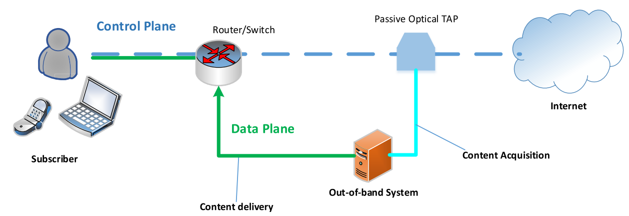


FIGURE 4. the Out-of-band architecture.

SPAN ports. Qwilt QB-Series is an example of out-of-band transparent caching system [34]. As shown in Fig. 4, in this system, the control plane and data plane are separated. In the control plane, the cache manager inspects the request data traffic from subscriber using deep packet inspection(DPI), policy-based routing(PBR), and the border gateway protocol(BGP) and then diverts packets to the data plane cache servers. In the data plane, the cache server stores and delivers the most popular content to improve the QoE for the end-users.

Transparent caching is a kind of partial solution to cope with the explosive growth of Internet traffic, especially the over-the-top (OTT) data traffic. It can enable ISP to improve the efficiency of content delivery and ease the network congestion, thus improving the QoE of end-users. Moreover, the transparent caching can also reduce the network bandwidth consumption, capital expense (CAPEX) and operating expense (OPEX). However, transparent caching also has some challenges, such as the encrypted content delivery. In conclusion, transparent caching is an effective collaboration scheme for content distribution and network infrastructures.

F. THE COOPERATIVE CACHING

With the explosive growth of network traffic, especially the P2P traffic and online video traffic, deploying a large number of content caches has been the significant method to reduce the transit traffic costs of ISPs. Meanwhile, how to optimize these distributed caches also has become the important research issue. In recent years, the cooperative caching has been considered as an extremely valuable approach and gained a lot of attention from the academia and industry. We consider the collaborative caching is a significant collaboration solution for content delivery and network infrastructures. In this subsection, we mainly introduce the cooperative caching solutions in wired network and wireless network.

In the wired network, some cooperative caching models or solutions are proposed. In [35], Dán propose a cooperative caching model based on game-theoretic approach, which models the network of cooperating caches as an n-person noncooperative game. Dai *et al.* in [36] propose a collaborative caching solution in ISP-aware P2P networks aimed to minimize costly inter-ISP traffic. And they develop an inter-ISP traffic model by taking P2P video streaming as a representative application, and according to this model they propose a optimization framework to address challenges in allocating resources on collaborative cache servers. Moreover, Fortino *et al.* in [37] propose a CDN-based distributed multimedia system called COMODIN, which can provide interactive and collaborative multimedia services to a group of users. The COMODIN can not only improve the media content delivery, but also can increase the performance of the streaming control. Based on the COMODIN framework, G. Fortino *et al.* further propose a hierarchical control protocol for group-oriented playbacks in [38], which allows a synchronous group of users to share the control of the media streaming session provided by a Content Distribution Network.

With the rapid development of wireless network technology, improving the efficiency of content delivery in wireless network becomes a significant research topic. In addition, deploying caching in wireless network has been considered as one of most important ways to reduce the delay, offload the backhaul traffic and improve the efficiency of content delivery. In [61], Niesen *et al.* consider the general caching problem from the perspective of information theory, and then a communication scheme for the caching problem is proposed, which can solve the problems of optimal cache selection, load balancing, and interference and noise mitigation separately, suggesting a layered approach. In [62], Golrezaei *et al.* propose a distributed cache scheme to solve the explosive increase of video traffic in wireless network [63]. In this proposed scheme, the base station called *helper* integrates the content storage function, which can cache the most popular video files, and transmit them, upon requests, via short-range wireless links to the user terminals. In [64], Bastug *et al.* discuss the limitations of current wireless networks and propose a novel proactive caching scheme deploying on the edge of 5G wireless network. In this proposed scheme, the proactive caching plays a crucial role, which can alleviate backhaul congestion. In [65], Wang *et al.* investigate the potential techniques for caching in 5G mobile networks, including evolved packet core (EPC) network caching and radio access network (RAN) caching, then a caching scheme based on content-centric networking is proposed, and the deployment of in-network caching into mobile networks is designed that can potentially help reduce user content access delay and mobile traffic load is demonstrated. In [66], Paul *et al.* present a novel ''cache-and-forward'' protocol architecture for mobile content delivery services in the future Internet. And this architecture can be implemented as an overlay on top of the Internet Protocol (IP), or as a clean slate protocol for next-generation networks.

After deploying the cache in wireless network, enhancing the collaboration for content delivery and wireless network infrastructures is necessary and important to improve the performance of content delivery. In [39], Pantisano *et al.* propose a collaborative content caching and content delivery framework in small cell network. And in this proposed solution, the small cell base stations (SBSs) can access the caches from the neighboring SBSs within the same

TABLE 2. Overview of revolutionary collaboration solutions.

Collaboration	Solutions	References	Contributions
The SDN-based collaboration	SDN-based CDN-ISP collaboration	Wichtlhuber *et al.* [68]	Implementing a fine grained, integrated traffic engineering for CDN traffic in the ISPs network based on SDN rather than on DNS.
	NetSoft	Wang *et al.* [69]	Proposing a software defined decentralized mobile network architecture toward 5G.
	SoftAir	Akyildiz *et al.* [70]	Proposing a software defined architecture for next generation (5G) wireless systems
The ICN-based collaboration	IICN	Lee *et al.* [71]	Deploying CDN over ICN
	age-based collaborative caching	Ming *et al.* [72]	Proposing a age-based collaborative caching
	WAVE	Cho *et al.* [73]	Presenting a popularity-based collaborative caching scheme
	CRCache	Wang *et al.* [74]	Proposing a new cross-layer collaboration scheme to optimize the cache utilization
SDN-ICN	SDN-based ICN	Veltri *et al.* [75], Syrivelis *et al.* [76], Vahlenkamp *et al.* [77], S. Salsano *et al.* [78]	Proposing a SDN-based ICN architecture
	C-flow	Chang *et al.* [79]	An SDN-based content delivery framework, which supports name-based routing and caching, and can take advantage of existing traditional IP networks
	Content-based traffic engineering	Chanda *et al.* [80]	Proposing software defined information centric network architecture, and a solution of optimizing content delivery in this architecture.
	SDN-Based Autonomic CCN Traffic Management	Q. Sun *et al.* [81]	Solving the traffic management regarding the flow awareness and fairness as well as node-to-node collaboration

network domain, not just the core network. In addition, Chen *et al.* [40] consider the tradeoff between transmission diversity and content diversity, and then propose a SBS cooperation schemes in cluster-centric cache-enabled small cell networks. In [41], Khreishah and Chakareski propose a collaborative caching scheme in cellular networks, which transforms the collaborative caching problem into the optimization problem. And they divide data traffic into two categories: coded data and no-coded data. In the non-coded case, each content item is either cached completely at a base station or not at all, and they formulate this problem as an integer programming. And in the coded case, the coded content item can be stored at multiple base stations, and they formulate this problem as a linear program. In [42], Dai *et al.* propose a collaborative caching mechanism in wireless network based on Vickrey-Clarke-Groves (VCG) auction theory [67]. Specially, they focus on engineering the incentives to promote and encourage the cache servers owned by different wireless service providers (WSPs) to truthfully cooperate with one another. In addition, some works of the cooperative caching are studied from the perspective of energy efficiency. For example, in [43], Xu *et al.* build an energy consumption model, and then propose a practical scheme for coordinating the content placement and request routing to minimize the energy consumption of eNodeB caches.

Moreover, the collaboration solution for CDN and mobile networks also has got a lot of attention. In [44], Loulloudes *et al.* introduce the mobile content delivery networks (CDNs), and investigate how to improve the information dissemination by mobile CDNs. In [45], Yousaf *et al.* present a deployment solution of mobile CDNs in mobile network, and then propose a framework for fair video delivery to improve the QoE of end-users. And in [46], Munaretto *et al.* propose a content delivery framework that can retrieves the video content requested by the users from the local CDN cache, according to the mobility of the users.

The cooperative caching is a significant cache optimization method, which plays an important role in the collaboration solution for content delivery and network infrastructures. The cooperative caching can optimize the cache resource allocation and scheduling, and thus make content distribution more efficient. Although there are a lot of works on cooperative caching, they mainly focus on the theoretical study and simulation verification. Hence, how to deploy the cooperative caching in actual network is still an interesting and challenging work.

III. REVOLUTIONARY COLLABORATION FOR CONTENT DELIVERY AND NETWORK INFRASTRUCTURES

In recent years, the clean-slate approaches including SDN and ICN have become the research focus of academia and industry, which will bring great opportunities and challenges. These new approaches can make the network more flexible and controllable as well as make the content distribution more efficient. Hence, the clean-slate approaches can provide more possibilities and opportunities of collaboration for content delivery and network infrastructures. In this section, we first introduce the clean-slate network technologies, which include SDN and ICN. And then, we present the revolutionary collaboration solutions based on SDN and/or ICN. And these collaboration solutions are summarized in Table 2.

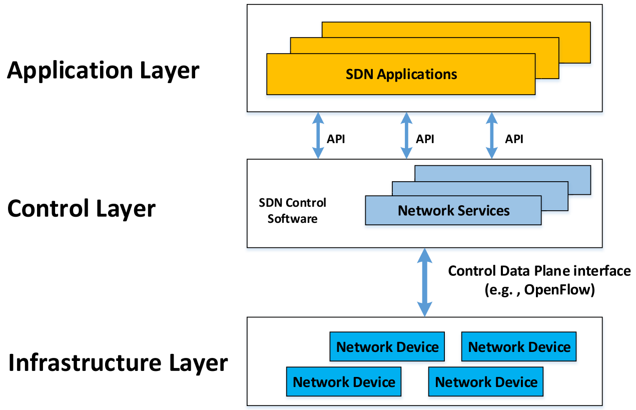


FIGURE 5. the SDN architecture [83].

A. SOFTWARE-DEFINED NETWORKING

Software-defined networking is an emerging network architecture that decouples the control plane from the data plane and provides programmability for network application development. These two features can bring great improvements and benefits to the current network. In SDN architecture, the network devices in the data plane such as routers, switches and so on, only focus on forwarding the data without considering the decision strategies. And all the decision strategies are decided and processed by the SDN controllers using the software programming in the control plane. So it brings great convenience for the network configuration and management. The network administrators can programmatically configure and manage the network in a centralized way, without requiring independently accessing and configuring each of the networks hardware devices [11], [82].

Fig. 5 depicts a reference model of the SDN architecture [83]. In this model, SDN architecture consists of three layers, namely infrastructure layer, control layer, and application layer. The infrastructure layer consists of data forwarding devices including routers and switches, which are responsible for collecting network status and processing data packets based on rules provided by a controller. The control layer mainly contains SDN controllers, bridging the application layer and the infrastructure layer. And network intelligence is (logically) centralized in software-based SDN controllers, which maintain a global view of the network. And the application layer contains SDN applications designed to fulfill user requirements [11]. Moreover, The interface between the layers is the one of the key techniques in SDN, which contains the south-bound interface and the north-bound interface.

In addition, Network Function Virtualization (NFV) is another important technology, which is closely coupled with SDN. It is necessary to briefly introduce NFV to better understand SDN-based collaboration solutions. The main idea of NFV is decoupling network functions from physical devices by leveraging virtualization technology. Thus, a given network service can be decomposed into a set of Virtual Network Functions (VNFs), which could be implemented in software running on one or more industry standard physical servers. And the VNFs may then be relocated and instantiated at different network locations without necessarily requiring the purchase and installation of new hardware [26].

B. INFORMATION-CENTRIC NETWORKING

Compared to the SDN, Information-centric networking (ICN) is a more revolutionary network architecture, which aims at achieving the content-centric communication instead of host-centric end-to-end communication. With the explosive growth of data traffic, ICN as a promising network technology has been increasingly attracting attentions from both academia and industry. Moreover, some ICN oriented research projects have gained a lot of great achievements, such as Data-Oriented Network Architecture (DONA) [84], Publish-Subscribe Internet Technology (PURSUIT) [85], Scalable and Adaptive Internet soLutions (SAIL) [86], content mediator architecture for content-aware networks (COMET) [87], Content Centric Networking (CCN) [88], Named Data Networking (NDN) [89], Mobility-First [90], etc. Although there are so many ICN research projects with different implementation methods, almost all of them focus on information naming, content delivery, mobility, security and so on [14]. Particularly, we mainly focus on content delivery characteristics of ICN in this paper. In this subsection, we briefly present NDN architecture.

NDN is a receiver-driven, content-centric novel network architecture. Content consumers perform communication through exchanging two types data packets: Interest Packets and Data Packets. And both types of packets carry a name that uniquely identifies a piece of data. In NDN architecture, the NDN router plays an important role, which maintains three data structures: a Pending Interest Table (PIT), a Forwarding Information Base (FIB), and a Content Store (CS). The FIB is responsible for forwarding Interest packets toward potential sources of matching Data. The CS is responsible for caching Data packets. And the PIT is responsible for keeping track of Interests forward upstream toward content sources, so that Data can be sent downstream to its requesters [91].

The NDN communication mechanism is described in Fig. 6. The content consumer sends a interest packet to the NDN network for requesting the required content, and when the Interest packet arrives at an NDN router, the NDN router first checks the CS for matching data. If the CS exists the targeted content, the router will return the Data packet on the interface from which the Interest came. Otherwise the router looks up the name in its PIT, and if a matching entry exists, it simply records the incoming interface of this Interest in the PIT entry. In the absence of a matching PIT entry, the router will forward the Interest toward the data producer(s) based on information in the FIB. When the a data arrives, an NDN router finds the matching PIT entry and forwards the data to all downstream interfaces listed in that PIT entry. Then it removes that PIT entry, and caches the Data in the Content Store.

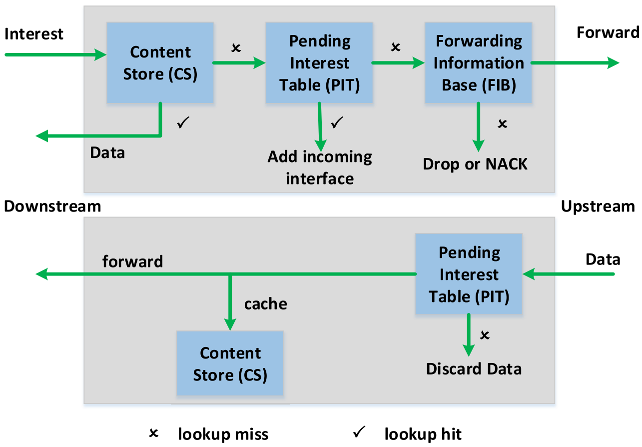


FIGURE 6. the NDN architecture [15].

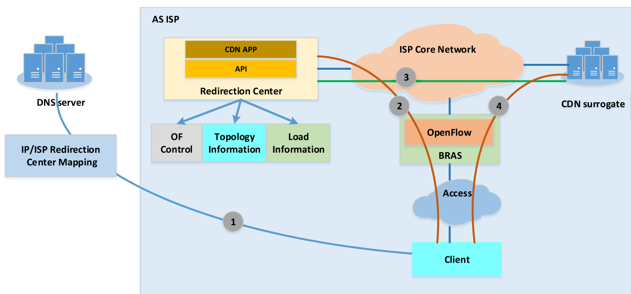


FIGURE 7. Overview of the System Design: (1) user resolves CDN URL via DNS and is redirected to the ISPs RC; (2) RC terminates HTTP session; (3) session is redirected to a suitable CDN surrogate using packet rewriting at the ingress router; (4) client is served from CDN surrogate [68].

C. THE SDN-BASED COLLABORATION SOLUTION

Because of the huge advantage in real-time centralized control and flexible network configuration as well as easy software implementation for new idea, SDN has been considered as the one of the most important network technology, especially in the aspect of content distribution. Accordingly, making full use of SDN to realize efficient content distribution is an important direction. And a lot of SDN-based collaborative content delivery solutions have been proposed [68]–[70]. In the following section, we present the SDN-based collaboration solutions.

The existing CDN-ISP collaboration solution is that ISP provides information of network status to help CDN improve its performance. In essence, the collaboration solution is based on DNS redirection. However, long lived, high-volume flows such as VoD traffic flows are very difficult to manage using the DNS approach. Because once a surrogate server is assigned, the content is usually delivered from the single server as a download of consecutive byte range requests using the same long lived, high-volume flow. This makes it difficult for CDN to deal with the flash crowds and link congestion. In order to solve this problem, an SDN-based CDN-ISP collaboration architecture is proposed [68], which is shown in Fig. 7. This architecture can implement a fine

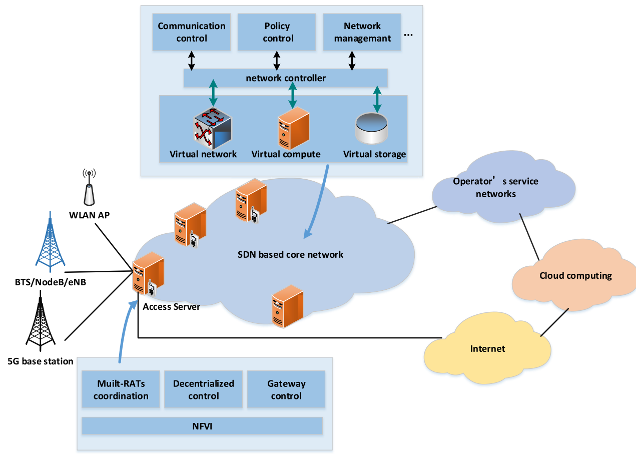


FIGURE 8. Architecture of SoftNet [69].

grained, integrated traffic engineering for CDN traffic in the ISPs network based on SDN rather than on DNS. Moreover, this architecture can implement the preservation of hidden information to incentivize the participating parties to apply this system. In addition, the power of the ultimate decision on the selection of surrogates belongs to CDN provider.

In addition to CDN, the collaboration between SDN and mobile network is also an important research direction. With the rapid development of mobile network technology, the demands for contents accessed by mobile clients are very high. Thus a lot of excellent works have been done, and they consider that the SDN-based collaboration approach can greatly improve the efficiency of content delivery and the QoE for users.

To cope with the challenges of mobile network, such as limited system capacity, high signaling overhead, inefficient data forwarding, high cost and poor scalability, etc. Wang *et al.* propose a software defined decentralized mobile network architecture toward 5G named NetSoft in [69], which is a typically SDN and mobile network collaboration paradigm and has advantages in content delivery. As shown in Fig. 8, this architecture consists of a unified radio access network (RAN) and an SDN-based core network, and all radio access points in unified RAN connect with access servers deployed at the edge of the SDN-based core network. In unified RAN, the multi-RATs (radio access technologies) coordination function in the access server can improve the content delivery by selecting the optimal RATs according to the wireless network conditions. Moreover, the mobile terminals served by the radio access points can either visit the operators service networks or third party service platforms, such as a cloud computing platform via the core network, or access the Internet or CDN server via a distributed gateway function within the access server. And gateway control function can select the optimal content delivery paths according to the traffic condition, thus improving the efficiency of content delivery.

In addition, Akyildiz *et al.* [70] also propose a novel software defined architecture for the 5G wireless network

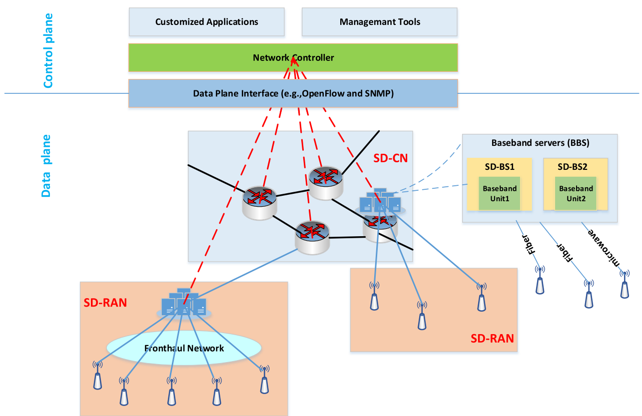


FIGURE 9. Architecture of SoftAir [70].

called SoftAir. As shown in Fig. 9, this architecture consists of a data plane and a control plane. The data plane is an open, programmable, and virtualizable network forwarding infrastructure, which consists of software-defined radio access network (SD-RAN) and software-defined core network (SD-CN). And the SD-RAN consists of a set of software-defined base stations (SD-BSs), while the SD-CN is composed of a collection of software-defined switches (SD-switches). The control plane mainly consists of two components: network management tools, and customized applications of service providers or virtual network operators. Moreover, in order to improve the efficiency of content delivery, the SoftAir adopts many methods to monitor network and optimize traffic control, such as mobility-aware control traffic balancing, distributed and collaborative traffic classifier. And the software-defined traffic engineering in SoftAir can be considered the collaboration solution for content delivery and network infrastructures.

Besides, there are still a lot of works focusing on the SDN-based collaboration solution. Nam *et al.* [92] leverage the SDN concept to implement intelligent content delivery. Li *et al.* [93] consider that SDN can simplify the design and management of cellular data networks, and propose a cellular SDN architecture, which offers fine-grain, real-time control without sacrificing scalability. In [94], Niephaus *et al.* also present a software defined wireless network framework, which aims at improving the efficiency content delivery in the wireless network. In [95], Chen *et al.* jointly consider networking, caching and computing techniques, and propose an integrated framework to improve end-to-end system performance. In addition, Huo *et al.* [96] expand the integrated framework to the wireless network, and propose a scheme to support energy-efficient information retrieval and computing services in green wireless networks.

SDN has great advantages in real-time network monitoring, configuration and control as well as deployment of new applications. By utilizing the controllability and flexibility of SDN, SDN-based collaboration solutions can greatly improve the efficiency of content delivery. In particular, with the continuous development of Internet of things and mobile

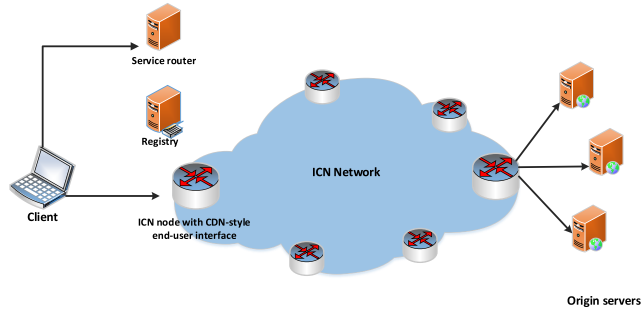


FIGURE 10. the IICN architecture.

Internet, the content delivery in the mobile network will face a lot of challenges. Hence, SDN-based mobile network is one of the most important research fields.

D. THE ICN-BASED COLLABORATION SOLUTION

ICN has emerged as a promising candidate for the future Internet architecture and it has a great advantage on the content delivery due to the efficient caching, content-oriented routing and forwarding. In order to distribute content more efficiently in ICN, it is necessary to enhance the collaboration in ICN, and the ICN-based collaboration schemes also have attracted more and more attention from the research community. In [71], a novel architecture called IICN has been proposed, which deploys CDN over ICN. As shown in Fig. 10, the IICN architecture consists of service router, ICN node and registry. In IICN, the ICN nodes have the CDN-style end-user interface and surrogates, but interact with each other using ICN technologies. Meanwhile, ICN nodes have a built-in reverse proxy functionality to interoperate with content origin servers. In addition, registry manages mappings between information identifiers and their locations. Moreover, the service router selects an ICN node to redirect the request of end-user.

The universal caching is a significant reason for the efficient content delivery of ICN [19]. And the collaborative caching in ICN is also considered extremely efficient and important. Ming *et al.* propose a age-based collaborative caching in [72], which implements a light-weight collaboration mechanisms by dynamically adjusting contents age at different network nodes. In [73], Cho *et al.* present a popularity-based collaborative caching scheme, which aims at getting efficient content delivery and cache usage while lowering the overhead of cache management. Moreover, Wang *et al.* propose a new cross-layer scheme named CRCache in [74], which exploits the correlation between content popularity and network information and selectively caches the most popular content on the best-matched routers.

ICN-based collaboration solution plays an important role in revolutionary collaboration solutions, which aims to solve the issues of efficient content delivery in ICN. In addition, in-network caching is the main reason for efficient content delivery, and collaborative caching is considered as an important caching optimization method in ICN. Accordingly, ICN-based caching solution has been considered as one of the most important collaboration solutions for content delivery and network infrastructures.

E. THE COLLABORATION SOLUTION FOR ICN AND SDN

Today, the services and applications based on content occupy a large proportion of network traffic, and the significance of content is becoming more and more prominent. Therefore, how to efficiently support content distribution has become an important topic in the research community. Obviously, ICN plays an important role in solving this problem. Meanwhile, SDN as a new kind of network technology can decouple the control plane from the data plane and provide programmability for network application development. Therefore, the ICN-SDN collaboration solution is a promising approach for the development of future network.

Recently, there exist some efforts to combine ICN with SDN. In [79], Chang *et al.* propose a novel network architecture named C-flow, an SDN-based content delivery framework, which supports name-based routing and caching, and can take advantage of existing traditional IP networks. This scheme addresses three basic problem: mapping content to IP address, name-based content delivery and name-based in-network storage. Moreover, in order to enhance the efficiency of content delivery, this scheme presents dynamic re-routing, parallel transmissions and in-memory cache management, which make full use of the controllability of SDN.

The idea of SDN centralized control can be used to content delivery in ICN, which can achieve efficient content delivery and traffic engineering. Chanda *et al.* [80] present a content centric network architecture using software defined networking principles, which implements efficient metadata driven services by observing and extracting content metadata at the network layer. In this scheme, both the content request from the client and the content publishing from the server are processed by controller in the control plane. Through centralized control based on SDN, the content management and content delivery can be optimized, which result in better utilization of network resources and QoE.

Observing that the current CCN traffic control is lacking of flow awareness, fairness and node-to-node collaboration, Sun *et al.* propose an SDN-based autonomic system as the control plane to supplement CCN traffic management in [81]. Correspondingly, CCN protocol and flow-aware content delivery as well as traffic control are supported by multi-protocol switches in the data plane. In particular, in order to implement the scheme, OF-Switch (OpenFlow-Switch) and CCN-Router cooperate as the data plane. This scheme integrates the advantages of CCN in content delivery and the advantages of SDN in fine-grained network control, thus extremely improving the efficiency and flexibility for the traffic management.

In [78], Salsano *et al.* propose and discuss solutions to support ICN by using SDN concepts. And they mainly focus on the ICN framework called CONET. In [75], Veltri *et al.* propose a collaboration framework, which deploys the coCONET (a ICN architecture) [97] in SDN. In addition, this scheme decouples the intelligence of coCONET from the forwarding (of interest packets and data packets) and caching functions. Besides, In [76], Syrivelis *et al.* also propose an OpenFlow-based framework, aiming at addressing the problem that how SDN and ICN could concretely be combined, deployed and tested. Moreover, Vahlenkamp *et al.* [77] propose a solution to enable information centric networking on existing IP networks, such as ISP or data center networks, using software-defined networking functions and control.

ICN has significant advantages in content delivery, and SDN has prominent advantages in controllability and flexibility. Hence, how to combine ICN and SDN becomes a significant issue. Although many excellent works have been done, the collaboration solutions for SDN and ICN still have a lot of challenging issues. Therefore, the collaboration for SDN and ICN is an important research direction in the future.

IV. COMPARISON AND ANALYSIS

In this section, we summarize, compare and analyze the advantages and disadvantages of these schemes from the perspective of technology, business and standardization.

A. FROM THE PERSPECTIVE OF TECHNOLOGY

This subsection, we compare and analyze evolutionary collaboration and revolutionary collaboration from the perspective of technology. Here, we mainly focus on the the performance of content delivery.

From the perspective of technology, the evolutionary collaboration solutions are the supplements to the traditional networks, especially for the CDN and P2P network. Hence, these solutions can be easily integrated with the current network, and the complexity is relatively low. Moreover, the evolutionary collaboration solutions have great advantages on content delivery and traffic control. For instance, compared with the native P2P, the P4P can reduce the average completion time approximately 20% in the real Internet experiments and can also significantly reduce the bottleneck link utilization [22]. In addition, the CaTE of [24] can reduce the link utilization up to 40%, and the network-wide traffic reduction can reach 15% in the peak hour. At the same time, the evolutionary collaboration solutions also have some insufficient aspects. For example, the P2P-based collaboration solutions mainly aim at popular content delivery, but it is inefficient for unpopular content delivery. The CDN-based collaboration solutions only provide service for a part of the subscribers rather than the all. The caching-based collaboration solutions have great advantages on technical complexity and deployment costs, but they still need to be improved in many ways. For example, how to deal with encrypted content.

Compared to the evolutionary collaboration solutions, the revolutionary collaboration solutions have great advantages on content delivery from the perspective of network architecture, because they inherently support content distribution. For example, the caching mechanism is one of the main

TABLE 3. The representative research and standard.

Solutions	Standards or Organization	Start time	goals
Evolutionary	IETF Application-Layer Traffic Optimization (ALTO)	2008	By enhancing the cognition of P2P to the network, improve the performance efficiency of a request/response, make the flow localization, reduce the pressure on the ISP network
	IETF Content Delivery Networks Interconnect (CDNI)	2011	Through collaboration between the CDN, as well as between CDN and ISP, implement across multiple CDN content distribution improve the efficiency of content delivery
Revolutionary	IRTF Information-Centric Networking Research Group (ICNRG)	2012	Guiding and promoting research of ICN
	IRTF Software-Defined Networking Research Group (SDNRG)	2013	Guiding and promoting research of SDN
	ONF OpenFlow	2008	Defining standard communications interface between the control and forwarding layers of SDN architecture
	CCNx	2009	Designing protocols to promote research of CCN

features of ICN, which can greatly improve the efficiency of content delivery. As for SDN, decoupling the control plane from the data plane can more easily control and optimize content delivery. Obviously, the revolutionary collaboration solutions are based on clean-slate approaches, which are greatly different from existing network architecture. Accordingly, there will be a lot of challenges to deploy the revolutionary collaboration solutions in the current network. Recently, there are a lot of works to compare CCN and CDN [98]–[101], which can be used as a reference for the comparison of revolutionary collaboration solutions and evolutionary collaboration solutions. The authors of [98] set up testbeds on the cloud computing platforms to compare the CDN and CCN, and the experiment results show that the network transmission performance of CDN is better than CCN in many case. In [99], the authors study and compare the performance of CCN and CDN, and the numerical results show that a CCN with small caches can provide significant performance gains compared to a traditional IP-based network. But a CDN with few replica servers can provide slightly better performance than a CCN in some scenarios, which is possible due to the fact that a CDN provides the additional degree of freedom to choose the location of the distributed cache. In fact, we should note that the revolutionary solutions are still in the initial stage. And there are many challenging problems to solve, such as content chunk decoding and restructuring, security and privacy, system reliability, etc. Therefore, it is necessary to further study the revolutionary solutions.

In short, the evolutionary collaboration solutions are very important to solve the problem of the content delivery for the current network, which can be quickly deployed. But revolutionary collaboration solutions can further improve the efficiency of content distribution.

B. FROM THE PERSPECTIVE OF BUSINESS

The comparison and analysis of business aspects are of extremely importance, because they are related to the commercialization of these collaboration solutions. In the aspect of business, we mainly focus on the modification cost or the investment cost for the current network.

Since the evolutionary collaboration solutions are based on the traditional content distribution technology, such as P2P, CDN. They have inherent advantages in deployment for the current network, which can make full use of the existing networks. Hence, The evolutionary collaboration solutions have lower modification costs compared to the revolutionary collaboration solutions. For instance, the P2P-ISP collaboration solutions only need to deploy a certain amount of servers in current network and provide some interfaces to communicate [9], [22]. As for the CDN-ISP collaboration solutions, deploying the systems inside the ISP network does not require large changes in the network configuration. In addition, the CaTE system even does not require any change in the network configuration or ISP DNS operation, which only requires the installation of one or more CaTE servers in an ISP and the establishment of a connection between ISP and CDN to facilitate communication between them [24].

The revolutionary collaboration solutions adopt some clean-slate approaches, which change a lot to the existing network infrastructure. Moreover, the content delivery devices need to support the SDN-based switches and/or content-based routers, which have greatly differences with the current network devices. As a result, the cost of deployment will be extremely high.

C. FROM THE PERSPECTIVE OF RESEARCH AND STANDARDIZATION

With the growth of Internet traffic, the collaboration for content delivery and network infrastructure has been considered more and more important. Moreover, the research of collaboration for content delivery and network infrastructure has attracted attention and participation from both academia and industry. In this subsection, we compare and analyze the collaboration schemes from the perspective of research and standardization. The representative researches and standards are summarized in TABLE 3.

The Internet Engineering Task Force (IETF) has done many works on evolutionary collaboration and published a lot of relevant standards and drafts. Application-Layer Traffic Optimization (ALTO) [102] is a significant working group in IETF, which aims to devise a request/response protocol for allowing a host to benefit from a server which is more cognizant of the network infrastructure than the host would be.

Besides, ALTO working group has developed an HTTP-based protocol to allow hosts to benefit from the network infrastructure by having access to a pair of maps: a topology map and a cost map. Therefore, this work can realize the performance optimization of overall network duo to the perception of the underlying network conditions. For instance, P2P peer can select the closer and better content nodes, CDN can select more optimal paths and content servers. In addition, Content Delivery Networks Interconnection (CDNI) [103] is another working group aiming to allow the interconnection of separately administered CDNs in support of the end-to-end delivery of content from Content Service Providers (CSPs) through multiple CDNs and ultimately to end users (via their respective User Agents). In essence, it is also a kind of collaboration which involves CDN and ISP.

Although the standards and drafts regarding revolutionary collaboration are relatively less, the revolutionary collaboration for content delivery and network infrastructure has attracted significant attention from both academia and industry. In particular, the current works mainly focus on SDN and ICN, and some standards and drafts have been published, such as OpenFlow [104], I2RS [105], CCNx [88], etc. Note that the revolutionary collaboration is at the initial stage of the study, especially ICN-based collaboration. In addition, the Internet Research Task Force (IRTF) has created Software-Defined Networking Research Group (SDNRG) [106] and Information-Centric Networking Research Group (ICNRG) [107] to guide and promote research of SDN and ICN, respectively.

In summary, the researches and standards regarding evolutionary collaboration are relatively mature due to the continuous improvement of the technology, such as P2P, CDN, etc. In contrast, the revolutionary collaboration is at incubation stages, and there are a lot of challenging issues that need to be addressed. Therefore, the evolutionary collaboration has advantages on practical deployment compared with the revolutionary collaboration. Nevertheless, the revolutionary collaboration can also bring great benefits for the network in the future. The comparison and analysis are summarized in TABLE 3.

V. CHALLENGES AND FUTURE DIRECTIONS

Although the research of collaboration for content delivery and network infrastructures has made great progress, it is still a new research area with a lot of challenging issues that need to be addressed. In this section, we present some important yet challenging problems, at the same time, we also outline possible future research directions.

A. NETWORK MONITORING

The network monitoring is one of the key techniques to realize collaboration for content delivery and network infrastructures. The goal of network monitoring is to provide real-time accurate information regarding the global network topology and link load condition for the control plane to make the optimal decision. In the evolutionary collaboration schemes, the network monitoring component can gather the detailed information about the network topology [24], [25]. For example, the Interior Gateway Protocol (IGP) listener can provide up-to-date information about routers and links. And the link utilization can be retrieved via Simple Network Management Protocol (SNMP) from the routers or an SNMP aggregator. Compared to the evolutionary collaboration schemes, the SDN-based collaboration schemes can collect network status via the SDN controllers, thus building a global view of the network.

Although a lot of works have been done on network monitoring, there are still a lot of challenging issues to be solved [108]. For example, the information obtained by network monitoring is often inaccurate due to the real-time change of the network state. A lot of network changes, such as snap up online, publishing popular video, usually lead to network traffic flash crowds. As a result, some optimal links may occur congestion and some content servers may became overload. Besides, the information obtained by network monitoring is often delayed due to the network process performance insufficient.

Recently, the big data has become an important technology, and combing with the big data method has been considered as an important research perspective for the network monitoring [109], [110]. The big data method for the collaboration includes real-time accurate measurement, perception, analysis and so on, however, how to realize intelligent efficient controlling and scheduling for network resource via the big data method still needs further research. In conclusion, real-time accurate network monitoring is still an important and interesting research direction.

B. NETWORK RELIABILITY AND SECURITY

Although revolutionary collaboration solutions can easily solve a lot of problems that the current network can hardly solve, such as centralized traffic scheduling and efficient content delivery, they have many problems on network reliability [111], [112]. Obviously, the network reliability is the base of collaboration for content delivery and network infrastructure. And it is also the significant precondition for network operators to practically deploy revolutionary collaboration solutions. For example, SDN can realize the flexible deployment and management for network applications and services via decoupling control plane from data plane and network programmability. However, this decoupled structure raises additional computational and network resources consumption that may even lead to fatal disasters [111]. In addition, ICN is still in the initial stage of exploration, and many research of ICN mainly focus on names, routing and forwarding and in-network caching [113]. Meanwhile, the reliability of ICN has not been explored much. Therefore, improving the network reliability is an important research direction, which should be paid more attention in the future.

Moreover, network security is a huge challenge for the revolutionary collaboration solutions. Some research works regarding the SDN security have been done. For example,

S. Scott-Hayward *et al.* present a comprehensive survey of the research relating to security in SDN in [114]. Sezer *et al.* present the security challenges of SDN in [115], including transport layer security (TLS), denial of service (DoS) attack, etc. In addition, Hakiri *et al.* in [116] also provide the challenges and opportunities for security in SDN. Besides, ICN also has a lot of security challenges, which includes trust model [117], privacy [118], [119], access control [120], [121] and attacks [122], etc. Although many works have been done to try to solve these problems, network security is still an important research direction in the future.

C. CONTENT DELIVERY IN MOBILE NETWORK

With the continuous development of mobile communication technology, more and more end users access the Internet via mobile devices, such as smartphone, tablet. As a result, mobile data traffic increases rapidly. According to [123], global mobile data traffic will increase 7-fold between 2014 and 2019. Mobile data traffic will grow at a CAGR of 47 percent between 2016 and 2021, reaching 49.0 exabytes per month by 2021. Therefore, content delivery in the mobile network is a challenging research subject.

Content delivery in mobile network still has at least two problems that need to be solved. The first is content deployment issue. Namely, where should the content be deployed, the base station, core network or other places? Recently, both academia and industry are paying substantial attention on the research of future mobile networks architecture and content delivery technology, such as mobile content delivery network (mCDN) [45], cloud radio access networks (C-RAN) [124], [125], mobile edge computing (MEC) [126], fog computing [127], caching in 5G mobile networks [65] etc. However, designing which kind of architecture to use and where to deploy the content are the extremely significant research direction. The second issue is how to efficiently collaborate between the content delivery and mobile network infrastructure. The goal is to realize the collaborative management between the content delivery and network resources via fully aware of the topology information and link load status of the mobile network. Therefore, How to improve the performance and efficiency of content delivery in the mobile network is an important research direction.

D. COMBINING WITH VIRTUALIZATION TECHNOLOGY AND CLOUD COMPUTING

The network cost and computing performance are important issues for the collaboration between content delivery and network infrastructures. In order to realize the low cost and high performance for the collaboration, combining with virtualization technology [26], [128] and cloud computing [129], [130] is an important yet challenging problem.

Recently, virtual content delivery network has been proposed [131], which deploys cache nodes on virtual machines, and provides storage resources, computing resources and network resources for content provider as required. In addition, network function virtualization built in 5G mobile wireless networks has been considered as the significant trend, which can flexibly deploy new business and optimize resource utilization [132]–[134]. Meanwhile, a lot of NFV-based architectures are proposed [135]–[137], which aims to make full use of the network hardware resources, make the operation and management for content delivery system more flexibly, so as to reduce OPEX and CAPEX. However, there are still many challenging problem that need to be addressed, such as system management, resource allocation, energy efficiency and so on [26].

In addition, cloud computing as a promising technique has achieved great success. With the continuous development of Internet, large capacity and high speed content delivery systems need to combine with cloud computing technique, which can provide efficient computing and storage. In [138], Zhu *et al.* study the cloud-aware multimedia and multimedia-aware cloud, and propose a media-edge cloud architecture to achieve a high QoS for multimedia services. Wen *et al.* [139] present a survey on cloud-based mobile media networks, and discuss the research efforts, including resource management and control, media platform services, cloud systems and applications etc. Combing cloud computing technology, the collaboration solutions can allocate resource on demand and schedule content more agilely, thus making the network more elastic and cost-effective. Therefore, combining with virtualization technology and cloud computing is an important research direction in the collaboration for content delivery and network infrastructures.

VI. CONCLUSION

This paper presents a survey of collaboration for content delivery and network infrastructures. We first introduce the evolutionary collaboration solutions, which include P2P-ISP, CDN-ISP, CDN-P2P, CDN-P2P-GRID-Agent, transparent caching and collaborative caching. Then, the emerging network techniques, including SDN and ICN, are presented. And we also discuss the revolutionary collaboration solutions which have adopted these new network technologies. Besides, we compare the evolutionary solutions and revolutionary solutions from the perspective of technology, business and standardization. And we analyze the advantages and disadvantages among the collaboration solutions. Finally, we present the challenges and future directions. In summary, the research on the collaboration for content delivery and network infrastructures is of great value, and there are a lot of issues and challenges that need to be addressed. It is necessary to take much more efforts to develop more efficient, flexible and practical collaboration methods, so as to greatly improve the content delivery in the future.

ACKNOWLEDGMENT

The authors thank the reviewers for their detailed reviews and constructive comments, which have helped to improve the quality of this paper.

REFERENCES

[1] "Cisco visual networking index: Forecast and methodology, 2015—2020," Cisco, San Francisco, CA, USA, White Paper, 2016. [Online] Available: https://www.cisco.com/c/dam/en/us/solutions/collateral/service-provider/global-cloud-index-gci/white-paper-c11-738085.pdf

[2] J. Brutlag, "Speed matters for Google Web search," Google, Inc., Mountain View, CA, USA, Tech. Rep., Jun. 2009. [Online] Available: https://research.googleblog.com/2009/06/speed-matters.html

[3] T. Flach *et al.*, "Reducing Web latency: The virtue of gentle aggression," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 4, pp. 159–170, 2013.

[4] Amazon. (2008). *Insights Into in-Memory Computing and Real-Time Analytics*. [Online] Available: http://blog.gigaspaces.com/amazon-found-every-100ms-of-latency-cost-them-1-in-sales/

[5] S. Androutsellis-Theotokis and D. Spinellis, "A survey of peer-to-peer content distribution technologies," *ACM Comput. Surv.*, vol. 36, no. 4, pp. 335–371, 2004.

[6] B. Pourebrahimi, S. Vassiliadis, and K. Bertels, "A survey of peer-to-peer networks," in *Proc. 16th Annu. Workshop Circuits, Syst. Signal Process. (ProRisc)*, 2005, pp. 263–270.

[7] A.-M. K. Pathan and R. Buyya, "A taxonomy and survey of content delivery networks," Grid Comput. Distrib Syst. Lab, Univ. Melbourne, VIC, Australia, Tech. Rep. GRIDS-TR-2007-4, 2007.

[8] G. Pallis and A. Vakali, "Insight and perspectives for content delivery networks," *Commun. ACM*, vol. 49, no. 1, pp. 101–106, 2006.

[9] V. Aggarwal, A. Feldmann, and C. Scheideler, "Can ISPS and P2P users cooperate for improved performance?" *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 3, pp. 29–40, 2007.

[10] I. Poese, B. Frank, B. Ager, G. Smaragdakis, and A. Feldmann, "Improving content delivery using provider-aided distance information," in *Proc. 10th ACM SIGCOMM Conf. Internet Meas. (IMC)*, 2010, pp. 22–34.

[11] W. Xia, Y. Wen, C. H. Foh, D. Niyato, and H. Xie, "A survey on software-defined networking," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 1, pp. 27–51, 1st Quart., 2014.

[12] B. A. A. Nunes, M. Mendonca, X.-N. Nguyen, K. Obraczka, and T. Turletti, "A survey of software-defined networking: Past, present, and future of programmable networks," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 3, pp. 1617–1634, 3rd Quart., 2014.

[13] T. Huang, F. R. Yu, C. Zhang, J. Liu, J. Zhang, and J. Liu, "A survey on large-scale software defined networking (SDN) testbeds: Approaches and challenges," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 2, pp. 891–917, 2nd Quart., 2016.

[14] G. Xylomenos *et al.*, "A survey of information-centric networking research," *IEEE Commun. Surveys Tuts.*, vol. 16, no. 2, pp. 1024–1049, May 2014.

[15] L. Zhang *et al.*, "Named data networking," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 44, no. 3, pp. 66–73, 2014.

[16] A. Passarella, "A survey on content-centric technologies for the current Internet: CDN and P2P solutions," *Comput. Commun.*, vol. 35, no. 1, pp. 1–32, 2012.

[17] Z. Lu, Y. Wang, and Y. R. Yang, "An analysis and comparison of CDN-P2P-hybrid content delivery system and model," *J. Commun.*, vol. 7, no. 3, pp. 232–245, 2012.

[18] B. Frank *et al.*, "Collaboration opportunities for content delivery and network infrastructures," *ACM SIGCOMM ebook Recent Adv. Netw.*, vol. 1, no. 1, pp. 305–377, Aug. 2013.

[19] G. Zhang, Y. Li, and T. Lin, "Caching in information centric networking: A survey," *Comput. Netw.*, vol. 57, no. 16, pp. 3128–3141, 2013.

[20] M. Zhang, H. Luo, and H. Zhang, "A survey of caching mechanisms in information-centric networking," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1473–1499, 3rd Quart., 2015.

[21] C. Fang, F. R. Yu, T. Huang, J. Liu, and Y. Liu, "A survey of energy-efficient caching in information-centric networking," *IEEE Commun. Mag.*, vol. 52, no. 11, pp. 122–129, Nov. 2014.

[22] H. Xie, Y. R. Yang, A. Krishnamurthy, Y. G. Liu, and A. Silberschatz, "P4P: Provider portal for applications," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 351–362, 2008.

[23] I. Poese, B. Frank, B. Ager, G. Smaragdakis, S. Uhlig, and A. Feldmann, "Improving content delivery with PaDIS," *IEEE Internet Comput.*, vol. 16, no. 3, pp. 46–52, May 2012.

[24] I. Poese, B. Frank, G. Smaragdakis, S. Uhlig, A. Feldmann, and B. Maggs, "Enabling content-aware traffic engineering," *Comput. Commun. Rev.*, vol. 42, no. 5, pp. 21–28, 2012.

[25] B. Frank *et al.*, "Pushing CDN-ISP collaboration to the limit," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 3, pp. 34–44, 2013.

[26] R. Mijumbi, J. Serrat, J.-L. Gorricho, N. Bouten, F. De Turck, and R. Boutaba, "Network function virtualization: State-of-the-art and research challenges," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 1, pp. 236–262, 1st Quart., 2015.

[27] J. Wu, Z. Lu, B. Liu, and S. Zhang, "PeerCDN: A novel P2P network assisted streaming content delivery network scheme," in *Proc. 8th IEEE Int. Conf. Comput. Inf. Technol. (CIT)*, Jul. 2008, pp. 601–606.

[28] H. Jiang, J. Li, Z. Li, and X. Bai, "Efficient large-scale content distribution with combination of CDN and P2P networks," *Int. J. Hybrid Inf. Technol.*, vol. 2, no. 2, pp. 13–24, 2009.

[29] D. R. Choffnes and F. E. Bustamante, "Taming the torrent: A practical approach to reducing cross-ISP traffic in peer-to-peer systems," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 38, no. 4, pp. 363–374, 2008.

[30] G. Fortino and W. Russo, "Using P2P, GRID and Agent technologies for the development of content distribution networks," *Future Generat. Comput. Syst.*, vol. 24, no. 3, pp. 180–190, 2008.

[31] G. Fortino and C. Mastroianni, "Editorial: Next generation content networks," *J. Netw. Comput. Appl.*, vol. 32, no. 5, pp. 941–942, Sep. 2009. [Online] Available: http://dx.doi.org/10.1016/j.jnca.2009.05.001

[32] G. Fortino and C. Mastroianni, "Special section: Content management and delivery through P2P-based content networks," *Multiagent Grid Syst.*, vol. 5, no. 2, pp. 133–135, 2009.

[33] White Paper. (2015). *PeerApp: Transparent Caching Primer*. [Online]. Available: http://www.peerapp.com/content-caching/

[34] C. Dixon, "Handling the explosion of online video: Why caching is the key to containing costs," nScreenMedia eBook, Oct. 2013. [Online] Available: http://qwilt.com/downloads/nscreenmedia-caching4scale.pdf

[35] G. Dán, "Cache-to-cache: Could ISPs cooperate to decrease peer-to-peer content distribution costs?" *IEEE Trans. Parallel Distrib. Syst.*, vol. 22, no. 9, pp. 1469–1482, Sep. 2011.

[36] J. Dai, B. Li, F. Liu, B. Li, and H. Jin, "On the efficiency of collaborative caching in ISP-aware P2P networks," in *Proc. IEEE INFOCOM*, Apr. 2011, pp. 1224–1232.

[37] G. Fortino, W. Russo, C. Mastroianni, C. E. Palau, and M. Esteve, "CDN-supported collaborative media streaming control," *IEEE Multimedia-Mag.*, vol. 14, no. 2, pp. 60–71, Apr. 2007.

[38] G. Fortino, C. Mastroianni, and W. Russo, "A hierarchical control protocol for group-oriented playbacks supported by content distribution networks," *J. Netw. Comput. Appl.*, vol. 32, no. 1, pp. 135–157, 2009.

[39] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "In-network caching and content placement in cooperative small cell networks," in *Proc. 1st Int. Conf. 5G Ubiquitous Connectivity*, Nov. 2014, pp. 128–133.

[40] Z. Chen, J. Lee, T. Q. S. Quek, and M. Kountouris, "Cluster-centric cache utilization design in cooperative small cell networks," in *Proc. IEEE ICC*, May 2016, pp. 1–6.

[41] A. Khreishah and J. Chakareski, "Collaborative caching for multicell-coordinated systems," in *Proc. IEEE INFOCOM*, Apr./May 2015, pp. 257–262.

[42] J. Dai, F. Liu, B. Li, B. Li, and J. Liu, "Collaborative caching in wireless video streaming through resource auctions," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 2, pp. 458–466, Feb. 2012.

[43] Y. Xu, Y. Li, Z. Wang, T. Lin, G. Zhang, and S. Ci, "Coordinated caching model for minimizing energy consumption in radio access network," in *Proc. IEEE ICC*, Jun. 2014, pp. 2406–2411.

[44] N. Loulloudes, G. Pallis, and M. D. Dikaiakos, "Information dissemination in mobile CDNs," in *Content Delivery Networks* (Lecture Notes Electrical Engineering), vol. 9. Berlin, Germany: Springer, 2007, pp. 343–366. [Online]. Available: https://doi.org/10.1007/978-3-540-77887-5_14

[45] F. Z. Yousaf, M. Liebsch, A. Maeder, and S. Schmid, "Mobile CDN enhancements for QoE-improved content delivery in mobile operator networks," *IEEE Netw.*, vol. 27, no. 2, pp. 14–21, Mar. 2013.

[46] D. Munaretto, F. Giust, G. Kunzmann, and M. Zorzi, "Performance analysis of dynamic adaptive video streaming over mobile content delivery networks," in *Proc. IEEE ICC*, Jun. 2014, pp. 1053–1058.

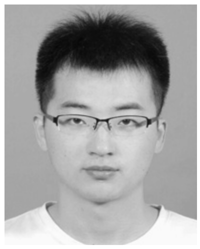
[47] N. M. M. K. Chowdhury and R. Boutaba, "A survey of network virtualization," *Comput. Netw.*, vol. 54, no. 5, pp. 862–876, 2010.

[48] N. M. M. K. Chowdhury and R. Boutaba, "Network virtualization: State of the art and research challenges," *IEEE Commun. Mag.*, vol. 47, no. 7, pp. 20–26, Jul. 2009.

[49] J. E. Smith and R. Nair, "The architecture of virtual machines," *Computer*, vol. 38, no. 5, pp. 32–38, May 2005.

[50] H. Jiang, J. Li, Z. Li, and J. Liu, "Efficient hierarchical content distribution using P2P technology," in *Proc. 16th IEEE Int. Conf. Netw. (ICON)*, Dec. 2008, pp. 1–6.
[51] Z. Chen, H. Yin, C. Lin, X. Liu, and Y. Chen, "Towards a trustworthy and controllable peer-server-peer media streaming: An analytical study and an industrial perspective," in *Proc. IEEE GLOBECOM*, Nov. 2007, pp. 2086–2090.
[52] D. Xu, S. S. Kulkarni, C. Rosenberg, and H.-K. Chai, "Analysis of a CDN–P2P hybrid architecture for cost-effective streaming media distribution," *Multimedia Syst.*, vol. 11, no. 4, pp. 383–399, 2006.
[53] H. Yin *et al.*, "Design and deployment of a hybrid CDN-P2P system for live video streaming: Experiences with LiveSky," in *Proc. ACM Multimedia Conf.*, 2009, pp. 25–34.
[54] H. Yin *et al.*, "LiveSky: Enhancing CDN with P2P," *ACM Trans. Multimedia Comput., Commun., Appl.*, vol. 6, no. 3, p. 16, 2010.
[55] X. Liu, H. Yin, C. Lin, Y. Liu, Z. Chen, and X. Xiao, "Performance analysis and industrial practice of peer-assisted content distribution network for large-scale live video streaming," in *Proc. 22nd Int. Conf. Adv. Inf. Netw. Appl.*, 2008, pp. 568–574.
[56] G. Fortino and C. Mastroianni, "Special section: Enhancing content networks with P2P, grid and agent technologies," *Future Generat. Comput. Syst.*, vol. 24, no. 3, pp. 177–179, 2008.
[57] G. Novelli, G. Pappalardo, C. Santoro, and E. Tramontana, "A grid-based infrastructure to support multimedia content distribution," in *Proc. Workshop P2P, Grid Agents Develop. Content Netw.*, 2007, pp. 57–64.
[58] M. Luck, P. McBurney, and C. Preist, "A manifesto for agent technology: Towards next generation computing," *Auto. Agents Multi-Agent Syst.*, vol. 9, no. 3, pp. 203–252, 2004.
[59] G. Fortino, W. Russo, and M. Vaccaro, "An agent-based approach for the design and analysis of content delivery networks," *J. Netw. Comput. Appl.*, vol. 37, no. 1, pp. 127–145, 2014.
[60] D. Ó. Coileáin and D. O'Mahony, "Accounting and accountability in content distribution architectures: A survey," *ACM Comput. Surv.*, vol. 47, no. 4, p. 59, 2015.
[61] U. Niesen, D. Shah, and G. W. Wornell, "Caching in wireless networks," *IEEE Trans. Inf. Theory*, vol. 58, no. 10, pp. 6524–6540, Oct. 2012.
[62] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 1107–1115.
[63] R. Xie, F. R. Yu, H. Ji, and Y. Li, "Energy-efficient resource allocation for heterogeneous cognitive radio networks with femtocells," *IEEE Trans. Wireless Commun.*, vol. 11, no. 11, pp. 3910–3920, Nov. 2012.
[64] E. Bastug, M. Bennis, and M. Debbah, "Living on the edge: The role of proactive caching in 5G wireless networks," *IEEE Commun. Mag.*, vol. 52, no. 8, pp. 82–89, Aug. 2014.
[65] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, "Cache in the air: Exploiting content caching and delivery techniques for 5G systems," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.
[66] S. Paul, R. Yates, D. Raychaudhuri, and J. Kurose, "The cache-and-forward network architecture for efficient mobile content delivery services in the future Internet," in *Proc. 1st ITU-T Kaleidoscope Acad. Conf. Innov. NGN, Future Netw. Services*, May 2008, pp. 367–374.
[67] N. Nisan and A. Ronen, "Computationally feasible VCG mechanisms," *J. Artif. Intell. Res.*, vol. 29, pp. 19–47, May 2007.
[68] M. Wichtlhuber, R. Reinecke, and D. Hausheer, "An SDN-based CDN/ISP collaboration architecture for managing high-volume flows," *IEEE Trans. Netw. Service Manage.*, vol. 12, no. 1, pp. 48–60, Mar. 2015.
[69] H. Wang, S. Chen, H. Xu, M. Ai, and Y. Shi, "SoftNet: A software defined decentralized mobile network architecture toward 5G," *IEEE Netw.*, vol. 29, no. 2, pp. 16–22, Mar. 2015.
[70] I. F. Akyildiz, P. Wang, and S.-C. Lin, "SoftAir: A software defined networking architecture for 5G wireless systems," *Comput. Netw.*, vol. 85, pp. 1–18, Jul. 2015.
[71] B. Lee, H. Jeon, S. Yoon, and H. Song, "Towards a CDN over ICN," in *Proc. Int. Conf. Data Commun. Netw.*, 2012, pp. 46–51.
[72] Z. Ming, M. Xu, and D. Wang, "Age-based cooperative caching in information-centric networking," in *Proc. IEEE INFOCOM*, Aug. 2012, pp. 268–273.
[73] K. Cho, M. Lee, K. Park, T. T. Kwon, Y. Choi, and S. Pack, "WAVE: Popularity-based and collaborative in-network caching for content-oriented networks," in *Proc. IEEE INFOCOM*, Mar. 2012, pp. 316–321.
[74] W. Wang *et al.*, "CRCache: Exploiting the correlation between content popularity and network topology information for ICN caching," in *Proc. IEEE ICC*, Jun. 2014, pp. 3191–3196.
[75] L. Veltri, G. Morabito, S. Salsano, N. Blefari-Melazzi, and A. Detti, "Supporting information-centric functionality in software defined networks," in *Proc. IEEE ICC*, Jun. 2012, pp. 6645–6650.
[76] D. Syrivelis *et al.*, "Pursuing a software defined information-centric network," in *Proc. Eur. Workshop Softw. Defined Netw. (EWSDN)*, 2012, pp. 103–108.
[77] M. Vahlenkamp, F. Schneider, D. Kutscher, and J. Seedorf, "Enabling information centric networking in IP networks using SDN," in *Proc. IEEE SDN Future Netw. Services (SDN4FNS)*, Nov. 2013, pp. 1–6.
[78] S. Salsano, N. Blefari-Melazzi, A. Detti, G. Morabito, and L. Veltri, "Information centric networking over SDN and OpenFlow: Architectural aspects and experiments on the OFELIA testbed," *Comput. Netw.*, vol. 57, no. 16, pp. 3207–3221, 2013.
[79] D. Chang, M. Kwak, N. Choi, T. Kwon, and Y. Choi, "C-flow: An efficient content delivery framework with OpenFlow," in *Proc. Int. Conf. Inf. Netw. (ICOIN)*, Feb. 2014, pp. 270–275.
[80] A. Chanda, C. Westphal, and D. Raychaudhuri, "Content based traffic engineering in software defined information centric networks," in *Proc. IEEE INFOCOM*, Apr. 2013, pp. 357–362.
[81] Q. Sun, W. Wang, Y. Hu, X. Que, and X. Gong, "SDN-based autonomic CCN traffic management," in *Proc. IEEE GLOBECOM*, Dec. 2014, pp. 183–187.
[82] C. J. Bernardos *et al.*, "An architecture for software defined wireless networking," *IEEE Wireless Commun.*, vol. 21, no. 3, pp. 52–61, Jun. 2014.
[83] "Software-defined networking: The new norm for networks," Open Netw. Found., ONF White Paper, vol. 2, pp. 2–6, 2012.
[84] T. Koponen *et al.*, "A data-oriented (and beyond) network architecture," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 37, no. 4, pp. 181–192, 2007.
[85] (2008). *FP7 PSIRP Project*. [Online]. Available: http://www.psirp.org/
[86] (2010). *FP7 SAIL Project*. [Online]. Available: http://www.sail-project.eu/
[87] (2010). *FP7 COMET Project*. [Online]. Available: http://www.comet-project.org/
[88] (2009). *PARC CCNx*. [Online]. Available: http://www.ccnx.org/specifications/
[89] (2010). *NSF NDN*. [Online]. Available: http://named-data.net/
[90] (2010). *NSF MobilityFirst*. [Online]. Available: http://mobilityfirst.winlab.rutgers.edu/
[91] A. Afanasyev *et al.*, "NFD developer's guide," Named Data Netw., San Diego, CA, USA, NDN Tech. Rep. NDN-0021, May 2015. [Online]. Available: http://named-data.net/techreports.html
[92] H. Nam, D. Calin, and H. Schulzrinne, "Intelligent content delivery over wireless via SDN," in *Proc. IEEE WCNC*, Mar. 2015, pp. 2185–2190.
[93] L. E. Li, Z. M. Mao, and J. Rexford, "Toward software-defined cellular networks," in *Proc. Eur. Workshop Softw. Defined Netw. (EWSDN)*, 2012, pp. 7–12.
[94] C. Niephaus, G. Ghinea, O. G. Aliu, S. Hadzic, and M. Kretschmer, "SDN in the wireless context—Towards full programmability of wireless network elements," in *Proc. 1st IEEE Conf. Netw. Softw. (NetSoft)*, Apr. 2015, pp. 1–6.
[95] Q. Chen, F. R. Yu, T. Huang, R. Xie, J. Liu, and Y. Liu, "An integrated framework for software defined networking, caching, and computing," *IEEE Netw.*, vol. 31, no. 3, pp. 46–55, May/Jun. 2017.
[96] R. Huo *et al.*, "Software defined networking, caching, and computing for green wireless networks," *IEEE Commun. Mag.*, vol. 54, no. 11, pp. 185–193, Nov. 2016.
[97] N. B. Melazzi, M. Cancellieri, A. Detti, M. Pomposini, and S. Salsano, "The CONET solution for information centric networking," Netw. Group, Univ. Rome, Rome, Italy, Tech. Rep. Version 0.4, 2012. [Online]. Available: http://netgroup.uniroma2.it/CONET/
[98] G. Ma and Z. Chen, "Comparative study on CCN and CDN," in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 169–170.
[99] M. Mangili, F. Martignon, and A. Capone, "A comparative study of content-centric and content-distribution networks: Performance and bounds," in *Proc. IEEE GLOBECOM*, Dec. 2013, pp. 1403–1409.
[100] A. Lertsinsrubtavee, P. Mekbungwan, and N. Weshsuwannarugs, "Comparing NDN and CDN performance for content distribution service in community wireless mesh network," in *Proc. Asian Internet Eng. Conf. (AINTEC)*, 2014, p. 43.
[101] M. Mangili, F. Martignon, and A. Capone, "Performance analysis of content-centric and content-delivery networks with evolving object popularity," *Comput. Netw.*, vol. 94, pp. 80–98, Jan. 2015.

[102] *IETF ALTO*. [Online]. Available: http://datatracker.ietf.org/wg/alto/documents/
[103] *IETF CDNI*. [Online]. Available: http://datatracker.ietf.org/wg/cdni/documents/
[104] *ONF OpenFlow*. [Online]. Available: https://www.opennetworking.org/
[105] *IETF I2RS*. [Online]. Available: https://datatracker.ietf.org/wg/i2rs/charter/
[106] *IRTF SDNRG*. [Online]. Available: https://irtf.org/sdnrg
[107] *IRTF ICNRG*. [Online]. Available: https://irtf.org/icnrg
[108] S. Lee, K. Levanti, and H. S. Kim, ''Network monitoring: Present and future,'' *Comput. Netw.*, vol. 65, pp. 84–98, Jun. 2014.
[109] A. Bar, A. Finamore, P. Casas, L. Golab, and M. Mellia, ''Large-scale network traffic monitoring with DBStream, a system for rolling big data analysis,'' in *Proc. IEEE Int. Conf. Big Data (Big Data)*, Oct. 2014, pp. 165–170.
[110] A. Bär, P. Casas, L. Golab, and A. Finamore, ''DBStream: An online aggregation, filtering and processing system for network traffic monitoring,'' in *Proc. IEEE Int. Wireless Commun. Mobile Comput. Conf. (IWCMC)*, Aug. 2014, pp. 611–616.
[111] X. Guan, B.-Y. Choi, and S. Song, ''Reliability and scalability issues in software defined network frameworks,'' in *Proc. IEEE 2nd GENI Res. Educ. Experim. (GREE)*, Mar. 2013, pp. 102–103.
[112] D. Kreutz, F. Ramos, and P. Verissimo, ''Towards secure and dependable software-defined networks,'' in *Proc. 2nd ACM SIGCOMM Workshop Hot Topics Softw. Defined Netw.*, 2013, pp. 55–60.
[113] B. Ahlgren, C. Dannewitz, C. Imbrenda, D. Kutscher, and B. Ohlman, ''A survey of information-centric networking,'' *IEEE Commun. Mag.*, vol. 50, no. 7, pp. 26–36, Jul. 2012.
[114] S. Scott-Hayward, G. O'Callaghan, and S. Sezer, ''SDN security: A survey,'' in *Proc. IEEE SDN Future Netw. Services (SDN4FNS)*, Nov. 2013, pp. 1–7.
[115] S. Sezer *et al.*, ''Are we ready for SDN? Implementation challenges for software-defined networks,'' *IEEE Commun. Mag.*, vol. 51, no. 7, pp. 36–43, Jul. 2013.
[116] A. Hakiri, A. Gokhale, P. Berthou, D. C. Schmidt, and T. Gayraud, ''Software-defined networking: Challenges and research opportunities for future Internet,'' *Comput. Netw.*, vol. 75, pp. 453–471, Dec. 2014.
[117] Y. Yu *et al.*, ''Schematizing trust in named data networking,'' in *Proc. 2nd Int. Conf. Inf.-Centric Netw.*, 2015, pp. 177–186.
[118] A. Chaabane, E. De Cristofaro, M. A. Kaafar, and E. Uzun, ''Privacy in content-oriented networking: Threats and countermeasures,'' *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 3, pp. 25–33, 2013.
[119] R. Tourani, S. Misra, J. Kliewer, S. Ortegel, and T. Mick, ''Catch me if you can: A practical framework to evade censorship in information-centric networks,'' in *Proc. 2nd Int. Conf. Inf.-Centric Netw.*, 2015, pp. 167–176.
[120] Q. Li, R. Sandhu, X. Zhang and M. Xu, ''Mandatory content access control for privacy protection in information centric networks,'' *IEEE Trans. Depend. Sec. Comput.*, vol. 14, no. 5, pp. 494–506, Sep. 2017.
[121] Q. Li, X. Zhang, Q. Zheng, R. Sandhu, and X. Fu, ''LIVE: Lightweight integrity verification and content access control for named data networking,'' *IEEE Trans. Inf. Forensics Security*, vol. 10, no. 2, pp. 308–320, Feb. 2015.
[122] E. G. AbdAllah, H. S. Hassanein, and M. Zulkernine, ''A survey of security attacks in information-centric networking,'' *IEEE Commun. Surveys Tuts.*, vol. 17, no. 3, pp. 1441–1454, 3rd Quart., 2015.
[123] ''Cisco visual networking index: Global mobile data traffic forecast update, 2016—2021,'' Cisco, San Francisco, CA, USA, White Paper, 2017. [Online] Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html
[124] ''C-RAN: The road towards green RAN,'' China Mobile Res. Inst., Beijing, China, White Paper VER3, vol. 2, 2013.
[125] J. Wu, Z. Zhang, Y. Hong, and Y. Wen, ''Cloud radio access network (C-RAN): A primer,'' *IEEE Netw.*, vol. 29, no. 1, pp. 35–41, Jan. 2015.
[126] ETSI. *Mobile Edge Computing*. [Online]. Available: http://www.etsi.org/technologies-clusters/technologies/mobile-edge-computing?highlight=YToxOntpOjA7czozOiJtZWMiO30=
[127] M. Peng and K. Zhang, ''Recent advances in fog radio access networks: Performance analysis and radio resource allocation,'' *IEEE Access*, vol. 4, pp. 5003–5009, 2016.
[128] *ETSI NFV*. Online]. Available: http://www.etsi.org/technologies-clusters/technologies/nfv
[129] P. M. Mell and T. Grance, ''The NIST definition of cloud computing,'' *Commun. ACM*, vol. 53, no. 6, p. 50, 2011.
[130] M. Armbrust *et al.*, ''A view of cloud computing,'' *Commun. ACM*, vol. 53, no. 4, pp. 50–58, 2010.
[131] T.-W. Um, H. Lee, W. Ryu, and J. K. Choi, ''Dynamic resource allocation and scheduling for cloud-based virtual content delivery networks,'' *ETRI J.*, vol. 36, no. 2, pp. 197–205, 2014.
[132] E. Hossain and M. Hasan, ''5G cellular: Key enabling technologies and research challenges,'' *IEEE Instrum. Meas. Mag.*, vol. 18, no. 3, pp. 11–21, Jun. 2015.
[133] S. Abdelwahab, B. Hamdaoui, M. Guizani, and T. Znati, ''Network function virtualization in 5G,'' *IEEE Commun. Mag.*, vol. 54, no. 4, pp. 84–91, Apr. 2016.
[134] Z. Feng, C. Qiu, Z. Feng, Z. Wei, W. Li, and P. Zhang, ''An effective approach to 5G: Wireless network virtualization,'' *IEEE Commun. Mag.*, vol. 53, no. 12, pp. 53–59, Dec. 2015.
[135] C. Liang, F. R. Yu, and X. Zhang, ''Information-centric network function virtualization over 5G mobile wireless networks,'' *IEEE Netw.*, vol. 29, no. 3, pp. 68–74, May/Jun. 2015.
[136] H. Hawilo, A. Shami, M. Mirahmadi, and R. Asal, ''NFV: State of the art, challenges, and implementation in next generation mobile networks (vEPC),'' *IEEE Netw.*, vol. 28, no. 6, pp. 18–26, Nov./Dec. 2014.
[137] J. Costa-Requena *et al.*, ''SDN and NFV integration in generalized mobile network architecture,'' in *Proc. Eur. Conf. Netw. Commun. (EuCNC)*, Jun./Jul. 2015, pp. 154–158.
[138] W. Zhu, C. Luo, J. Wang, and S. Li, ''Multimedia cloud computing,'' *IEEE Signal Process. Mag.*, vol. 28, no. 3, pp. 59–69, May 2011.
[139] Y. Wen, X. Zhu, J. J. P. C. Rodrigues, and C. W. Chen, ''Cloud mobile media: Reflections and outlook,'' *IEEE Trans. Multimedia*, vol. 16, no. 4, pp. 885–902, Jun. 2014.



QINGMIN JIA received the B.S. degree in communication engineering from Qingdao Technological University, Qingdao, China, in 2014. He is currently pursuing the Ph.D. degree with the State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications. His current research interests include future network architecture design, content delivery, network virtualization, and resource allocation.

RENCHAO XIE received the Ph.D. degree from the School of Information and Communication Engineering, Beijing University of Posts and Telecommunications, in 2012. From 2010 to 2011, he visited Carleton University, Ottawa, ON, Canada, as a Visiting Scholar. From 2012 to 2014, he held a post-doctoral position with China Unicom. He is currently an Associate Professor and a Master Supervisor with the Beijing University of Posts and Telecommunications. He has authored over 20 journal and conference papers. His current research interests include future network architecture design, information centric networking, and mobile content delivery network. He has served for several journals and conferences as a Reviewer, including the IEEE Transactions on Communications, the IEEE Transactions on Vehicular Technology, the *EURASIP Journal on Wireless Communications and Networking*, and the 2011 IEEE Global Communications Conference.

TAO HUANG received the B.S. degree in communication engineering from Nankai University, Tianjin, China, in 2002, and the M.S. and Ph.D. degrees in communication and information system from the Beijing University of Posts and Telecommunications, Beijing, China, in 2004 and 2007, respectively. He is currently a Professor with the Beijing University of Posts and Telecommunications. His current research interests include network architecture, routing and forwarding, and network virtualization.

JIANG LIU received the B.S. degree in electronics engineering from the Beijing Institute of Technology, Beijing, China, in 2005, the M.S. degree in communication and information system from Zhengzhou University, Zhengzhou, China, in 2009, and the Ph.D. degree from the Beijing University of Posts and Telecommunications, Beijing, in 2012. He is currently an Associate Professor with the Beijing University of Posts and Telecommunications. His current research interests include network architecture, network virtualization, software-defined networking, information centric networking, and tools and platforms for networking research and teaching.

YUNJIE LIU received the B.S. degree in technical physics from Peking University, Beijing, China, in 1968. He is currently the Academician of the China Academy of Engineering, the Chief of the Science and Technology Committee, China Unicom, and the Dean of the School of Information and Communications, Beijing University of Posts and Telecommunications. His current research interests include next generation network, network architecture, and management.