Software-Defined Live Video Delivery Networks with Device-to-Device Communications

Dianjie Lu[1, 2], Ren Han[3], Jie Tian[1, 2, 4], Weizhi Xu[1, 2], Xiangwei Zheng[1, 2], Guijuan Zhang[1, 2] , Hui Yu[1, 2]
[1]School of Information Science and Engineering, Shandong Normal University, Jinan, China, 250358
[2]Shandong Provincial Key Laboratory for Novel Distributed Computer Software Technology, Jinan, China, 250358
[3]School of Optical-Electrical and Computer Engineering, University of Shanghai for Science and Technology, Shanghai, China, 200093
[4]Shandong Provincial Key Laboratory of Computer Networks, Shandong Computer Science Center
(National Supercomputer Center in Jinan), Jinan, China, 250101
ludianjie@sina.com, campushr@163.com, tianjiesdu@gmail.com, xuweizhi@sdnu.edu.cn, sdnuzxw@126.com

ABSTRACT

Video Delivery Networks (VDNs) have been widely applied to the video content delivery of mobile Internet recently. However, the traditional VDNs are all infrastructure pre-deployed which makes them inflexible to adapt to the diversity and mass of live video delivery in mobile Internet. Thus, how to meet mobile users' live video demands while minimizing the delivery cost is still a challenge task.

To solve this problem, we propose a Software-defined live Video Delivery Network (S-VDN) framework with Device-to-Device (D2D) communication for mobile Internet. With a global view of the network, S-VDN is capable of managing the mobile users' requests and the network topology conveniently as well as scheduling the radio resources. In addition, S-VDN enables content delivery among closely located mobile users without going through base stations by employing the D2D technique. Furthermore, we formulate this optimization problem as a Multi-commodity Flow (MCF) problem and design two heuristic algorithms to obtain the optimal solution. Simulation experiments show that our proposed algorithms are efficient in terms of both performance and rental cost.

CCS CONCEPTS

• **Networks** → Mobile ad hoc networks

KEYWORDS

Video delivery networks, Software-defined networking, Device-to-Device communication, Multi-commodity flow

ACM Reference format:

D. Lu, R. Han, J. Tian, W. Xu, X. Zheng, G. Zhang, H. Yu. 2017. Software-Defined Live Video Delivery Networks with Device-to-Device Communications. In Proceedings of Mobimedia 2017, Chongqing, China,

Mobimedia 2017, July 13-14, Chongqing, People's Republic of China

reach 82% by 2020 [1]. Furthermore, with the development of mobile Internet, mobile users are becoming the dominant consumers of Internet videos which may enjoy or produce live videos whenever and wherever possible. For example, as the fourth largest Internet traffic producer, Twitch provide more than 150 billion minutes of live videos each month which are mainly generated by mobile users [2]. Thus, satisfying the explosively increasing demands of live videos for mobile users is a challenging task.

VDN services are obtaining increasing attention as the main video content delivery mode [3] [4]. However, the future development of VDNs faces some critical challenges. On the one hand, most VDNs rely on a pre-deployed infrastructure and statically configured network topology. On the other hand, the architectures are not suitable for numerous small video providers because they have no ability to deploy infrastructure which inevitably hinders the development of VDNs.

Mobile Internet regard the mobile network especially the cellular network as the access network through which users can surf the Internet anytime and anywhere using portable mobile terminals, such as smart phones, pads, and portable devices. In the cellular network, content providers can scale their video delivery services to the mobile users by taking advantage of the wireless broadband access. But the cellular network also faces the challenges in coping with the explosively growth of network traffic as the tremendous popularity of high data rate videos. First, the random requests of videos need a fast and efficient resoucre allocation. But the devices and protocols in current cellular networks cannot support this dynamic configuration. Second, the random mobility of users make the base stations hard to cover all of them because of the the wireless signal's poor coverage which will decrease the quality of service (QoS), such as the request acceptance ratio.

SDN has attracted much attention in recent years [5] [6]. Its core idea is to separate the control plane and data plane so as to realize the flexible control of network traffic. Via an SDN controller with a global view of the network, operators can manage and optimize resource allocation efficiently in response to time-varying network conditions. In addition, D2D technique can be employed in cellular networks which enables content delivery among closely located devices without going through base stations as in traditional cellular networks. In cellular networks, the tight coupling of femtocell and D2D communication has many advantages such as reducing the load traffic of base station and extending coverage[7][8].

In this paper, we propose a software-defined live video delivery network (S-VDN) framework with D2D communication, which can be appropriately applied to mobile Internet. Different from traditional VDNs, S-VDN is divided into two parts: the control plane and the data plane. In the control plane, the SDN controller with a global view of the VDN is capable of managing the mobile users' behaviors such as the requests and the topology as well as scheduling the radio resources. In the data plane, the video content is maintained on the origin servers and the users can forward the data quickly with D2D communication according to the scheduling results of the control plane. Furthermore, we formulate this optimization problem as a Multi-commodity Flow (MCF) problem and present two heuristic algorithms to obtain the optimal solution. S-VDN can also allocate and provide resources automatically according to the appearance and duration of dynamic videos. Experimental results show that the proposed algorithm can increase the revenue and the acceptance ratio while reducing the rental cost in the long run.

The remainder of this paper is organized as follows. Section 2 discusses related work. Section 3 introduces network models. Section 4 proposes the S-VDN framework. Section 5 presents the problem formulation. Section 6 shows the simulation results to evaluate our proposed algorithms. Section 7 gives the conclusions.

2 RELATED WORKS

Content delivery network (CDN) has always been one of the key technologies for content delivery on the Internet. Most researchers focus on the server selection or edge server selection problem [9-12] [4]. Zegura et al. [9] proposed a method that combined the server pushing and client detection technologies. Wendell et al. [10] proposed a distributed server selection mechanism that can effectively balance the load between the edge servers. The authors in [11] constructed a mobile mechanism to select edge servers using synchronized multimedia integration language (SMIL) files. In this mechanism, mobile terminals select appropriate surrogate sites by parsing the SMIL files that contain information on service content and the status of edge servers. ActiveCDN [12] allows content providers to designate particular servers dynamically according to user requests. Maggs et al. optimized the problem on matching clients and servers by formulating it as a stable marriage problem [4].

In addition, some efforts have been put into satisfying the QoS of terminal users. Rodolakis et al. [13] minimized the cost of storage and bandwidth under the constraint of delay. In [14], the authors investigated the QoS-aware replica placement problems in replica-aware and replica-blind scenarios. In [15], the authors proposed a hierarchical structure of mobile CDN to guarantee the QoS of the mobile terminal by selecting appropriate surrogate sites at different levels. In the literature [16], Xu et al. indicated that mapping the user requests to the proximal edge servers in the mobile environment can improve the QoS of the network significantly, such as delay and jitter. However, all of the CDNs above rely on a pre-deployed infrastructure.

Recently, the cloud CDN has been studied widely. Different to traditional CDNs, a cloud CDN can provide cost-saving services without owning infrastructure. Chen et al. [17] conducted the pioneer study on the replica placement problem in storage cloud CDN joint content delivery path construction. They formulated the problem as an integer-program problem and provided offline and online heuristics to solve it. Papagianni et al. [18] built a hierarchical framework for the CDN deployment problem on the multi-provider network cloud environment. In [19], the authors proposed a resource auto-scaling method to provide the bandwidth guarantees for video-on-demand applications. Hu et al. [20] studied the resource allocation and replica placement problems for cloud-based CDNs by considering diversities in user demand patterns. Mukerjee et al. [3] proposed a centralized VDN that allows network operators to control the video placement and bit-rate allocation at a fine timescale while optimizing the quality of live video delivery. However, they all assumed that contents are fixed and of the same size without considering the diversity of the live video delivery.

Overall, the traditional algorithms are inadaptable for delivery of live videos in mobile Internet because of the mobility of mobile users. The dynamics of the explosively increasing video contents aggravate the problem further, requiring new solutions to deliver live videos efficiently.

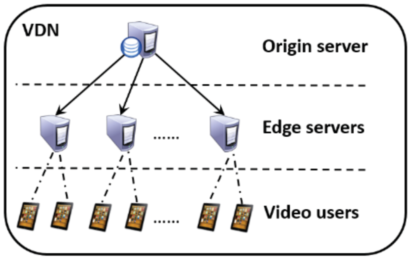


Fig. 1. Traditional VDN structure

3 NETWORK MODEL

Traditional VDNs generally consist of three kinds of roles [21] (shown in Fig. 1). Origin servers import videos into the networks. Edge servers replicate the videos from the origin server and serve the video users directly. In this situation, video contents should be pulled or pushed to the fixed edge servers from an origin server. It leads to low coverage and inflexibility of resource allocation which cannot be applied to the video delivery in Mobile Internet. In contrast, we propose a software-defined video delivery network model for Mobile Internet. In this model, the VDN can manage and optimize resource allocation efficiently in response to dynamic networks. In addition, it also uses the D2D technology which employs mobile users to delivery videos.

3.1 Configuration of the Software-defined VDN

We assume that a VDN can support one providers (or origin servers). In addition, a control center is provided to administer these providers. The origin servers can replicate the video content at multiple cloud sites (or surrogate sites) to serve all mobile user requests. We assume the existence of an origin server o , N

mobile users. The mobile users can be expressed by $\mathcal{N}=\{1,2,\ldots,n,\ldots,N\}$. Each mobile user in $\mathcal{N}$ has location coordinates (x_n, y_n). The VDN can be modeled as an undirected graph $\mathcal{G}=(\mathcal{H},\mathcal{E})$, where $\mathcal{H}\in\{o\cup\mathcal{N}\}$. Each link $e(i,j)\in\mathcal{E}$ between two mobile users i and j is associated with a bandwidth capacity weight value $B(e(i,j))$ (or B_{ij}).

Fig. 2 shows an example scenario of the VDN. The VDN is divided into two parts: the control plane and the data plane. In the control plane, the SDN controller with a global view of the VDN is capable of managing the mobile users' requests and scheduling the radio resources. In the data plane, the video content is maintained on the origin servers and the nodes forward the data quickly according to the scheduling results of the control plane. As we can see from Fig.2, the lines represent the potential delivery paths. Thereinto, the solid lines represent the communication to the base station and the dotted line represent the communication between mobile users. The base station covers the shadow area where the mobile users (such as n_1 - n_5) can get live video from the base station directly. In order to alleviate the traffic burden of the base station, parts of these mobile users (such as n_4, n_5) select the D2D communication to get the live video content. On the other hand, the mobile users out the coverage of the base station (such as n_6 - n_8) can only adopt the D2D communication.

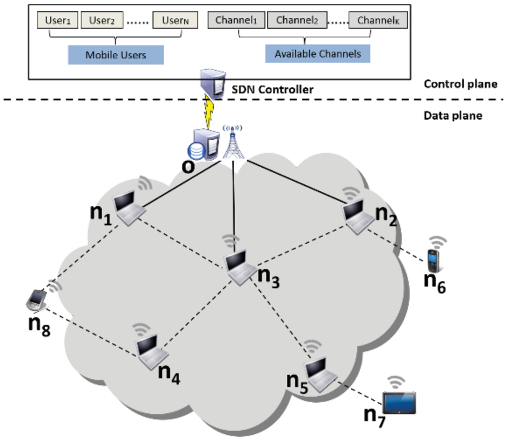


Fig. 2. Software-defined VDN model

3.2 Link Data Rate

In this network model, the mobile user can also be used as intermediate nodes for routing videos from the origin to other mobile users. We assume each mobile user uses 3G/4G/NxtG accessing technology and has two radio interfaces: one for communication with base stations (backhaul connections) and another for providing data forwarding to mobile users.

Supposing all mobile users exhibit the same power for transmission, the power spectral density from the transmitter is P_t. The widely used model [23] for power propagation gain is:

$$g_{ij}=\mu\cdot d_{ij}^{-n} \tag{1}$$

where n is the path loss factor, μ is an antenna-related constant, and d_{ij} is the distance between nodes i and j.

According to the Shannon–Hartley theorem, if node i sends data to node j on link $e(i,j)$ with the available bandwidth BW_{ij}, the data rate of the link is:

$$\mathcal{C}_{ij}=B_{ij}\log_2(1+\frac{g_{ij}P_t}{\eta}) \tag{2}$$

where η is the ambient Gaussian noise density.

4 SOFTWARE-DEFINED VDN FRAMEWORK

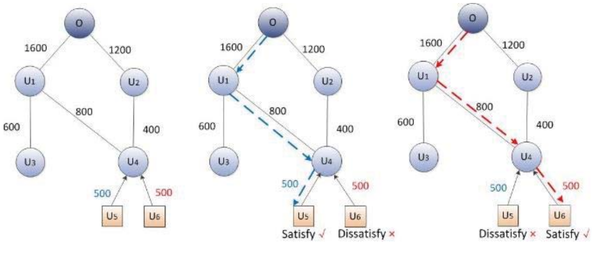
4.1 Session based Flow Scheduling

In order to avoid the frequent interaction between the data plane and the control plane, the coordinated rules are usually built based on the data flow, instead of data packet. To configure these flows fast and efficiently, we construct a session for each node pair from a mobile user to the origin server.

Under the S-VDN architecture, mobile users send their online video requests to nearby neighbor users. The SDN controller collects requests from different users and forms a set of video delivery sessions. Here, a set of $\mathcal{L}$ concurrent sessions is considered, and each of which is characterized by a node pair between the mobile user and an origin server with the rate requirement r_u. The session can be represented as $\ell(u,o)\in\mathcal{L}$, where $u\in\mathcal{U}$. Each session ℓ corresponds to a flow f in the network. Therefore, each flow starts from an origin server o and ends in a mobile user $u\in\mathcal{U}$, which is in contrast to the order of the request. The data rate demand for each session $\ell(u,o)$ is denoted as $r(\ell)$.

4.2 Path-splittable Scheduling

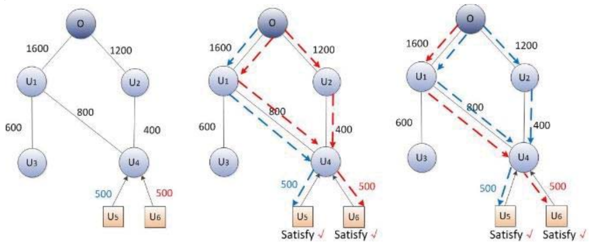
We assume that each mobile user can access multiple delivery delay node. If the VDN does not support the path-splitting, the performance of VDNs will decrease in acceptance rate and resource utilization rate [11] [12]. Fig. 3 (a) shows an example of VDN. The number beside the link is the bandwidth of this link. There are two mobile users $U5$ and $U6$ who send requests to the origin server o at the same time and the request of bandwidth is 500bps. Both of them are proximal to user $U4$. When we consider the link bandwidth capacity limitation, the bandwidth on the link $U2-U4$ is 400bps, which cannot satisfy the bandwidth of request. So, the delivery route has to select $U1-U4$. But the bandwidth of link $U1-U4$ is 800bps, which can only satisfy one request at most (shown in Fig. 3 (b) and (c)).



(a) (b) (c)

Fig.3 Path-Unsplittable Method

Fig. 4 (a) shows an S-VDN topology, which has the same network topology and users' request with Fig. 3 (a). The difference between the two methods is that path-splittable-supported method allows more links to provide bandwidth for each mobile user. As shown in Fig. 4 (b), the request of end user $U5$ is satisfied via the links $o-U1-U4-U5$. Because the link $U1-U4$ has been assigned to mobile user $U5$ with 500bps bandwidth, there is 300bps left. So, the request of $U6$ can be satisfied by employing two paths such as $o-U1-U4-U6$ and $o-U2-U4-U6$. On the other hand, we can also allocate one path $o-U1-U4-U6$ to $U6$ and two paths to $U5$ such as $o-U1-U4-U5$ and $o-U2-U4-U5$ (shown in Fig. 4 (c)).



(a) (b) (c)

Fig.4 Path-splittable Method

4.3 Spatial and Frequency Scheduling

The network directs an incoming request to a proximal mobile user that possesses sufficient bandwidth. We assume that each mobile user has an associated non-negative value D_u that expresses the physical range the user can cover. The request can only be directed to the mobile users in its coverage. We denote such a UE cluster by $\varphi(u)$ for each mobile user u:

$$\varphi(u)=\left\{n\in\mathcal{N}\,\middle|\,dis\left((x_u,y_u),(x_n,y_n)\right)\le D_u\right\} \qquad (3)$$

where $dis()$ is the function to compute the physical distance of two nodes.

4.4 Frequency Scheduling

Frequency scheduling can be conducted in frequency domain or time domain. In this paper, we only focus on frequency based channel assignment, i.e., how to assign channels at a mobile user for transmission and reception. In this paper, we adopt orthogonal subchannels for different users. We further let $\mathcal{K}=\{1,2,\ldots,k,\ldots,K\}$ denote the set of all subchannels in the system and $\mathcal{K}_i,\mathcal{K}_j\subseteq\mathcal{K}$ represent the set of available channels at user $i,j\in\mathcal{N}$. $\mathcal{K}_i$ may be different from $\mathcal{K}_j$, where j is not equal to i, and $j\in\mathcal{N}$, i.e., possibly $\mathcal{K}_i\ne\mathcal{K}_j$.

Assume channel k is available at both User i and User j, i.e., $k\in\mathcal{K}_i\cap\mathcal{K}_j$. We denote

$$w_{ij}^k=\begin{cases}1 & \text{If User } i \text{ transmit data to User } j \text{ on channel } k,\\ 0 & \text{otherwise.}\end{cases} \qquad (4)$$

For a UE $i\in\mathcal{N}$ and a channel $k\in\mathcal{K}_i$, denote φ_i^k the set of UEs that can also opportunistically access channel k and are within the transmission range of UE, i.e.,

$$\varphi_i^k=\{j\in\mathcal{N}\,|\,dis\left((x_i,y_i),(x_j,y_j)\right)\le D_u,k\in\mathcal{K}_j\} \qquad (5)$$

From the view of the transmitter, UE is not able to transmit to multiple UEs on the same channel. So, we can get

$$\sum_{j\in\varphi_i^k}w_{ij}^k\le 1 \qquad (6)$$

Thus, if node i sends data to node j on the available channel k with bandwidth B_k, the data rate of the link is:

$$\mathcal{C}_{ij}^k=w_{ij}^kB_k\log_2(1+\frac{g_{ij}P_t}{\eta}) \qquad (7)$$

When $w_{ij}^k=0$, we have $\mathcal{C}_{ij}^k=0$. According to the link data rate model, the aggregate flow rates on each link should never exceed the data rate of this link, which is an essential constraint for routing.

4.5 Routing Constraints based on the Sessions

Denoting $f_{ij}(\ell)$, the data rate on link (i,j) is attributed to session ℓ, where $i,j\in\mathcal{H}$. The traffic and storage request for each session ℓ is given by $r(\ell)$ and $c(\ell)$. The flow balance at each node i for each session ℓ can be described by the following equations. First, if the origin server i is the source node of session ℓ, that is, $i=s(\ell)$, then

$$\sum_{j\in\mathcal{H}}f_{ij}(\ell)-\sum_{m\in\mathcal{H}}f_{mi}(\ell)=r(\ell). \qquad (8)$$

Then, if node i is an intermediate relay node for session ℓ, that is, $i \neq s(\ell)$ and $i \neq d(\ell)$, then

$$\sum_{j\in\mathcal{H}} f_{ij}(\ell) - \sum_{m\in\mathcal{H}} f_{mi}(\ell) = 0 \quad (9)$$

Third, if mobile user i is the destination of session ℓ, that is, $i = d(\ell)$, then

$$\sum_{j\in\mathcal{H}} f_{ij}(\ell) - \sum_{m\in\mathcal{H}} f_{mi}(\ell) = -r(\ell) \quad (10)$$

Additionally, the aggregate flow rates of all channels on each link cannot exceed the capacity of this link. According to Equation (2), the constraint of the capacity of link (i, j) can be formulated as

$$\sum_{\ell\in\mathcal{L}} f_{ij}(\ell) \leq \sum_{k\in\mathcal{K}_i\cap\mathcal{K}_j} C_{ij}^k \quad (11)$$

5 PROBLEM FORMULATION

Mathematically, we formulate this optimization problem as a multi-commodity flow (MCF) problem.

$$Minimize \sum_{\ell\in\mathcal{L}} \sum_{e(i,j)\in\mathcal{E}} \sum_{k\in\mathcal{K}} \sigma_{ij} w_{ij}^k f_{ij}(\ell) \quad (12)$$

Subject to:

$$\sum_{j\in\mathcal{H}} f_{ij}(\ell) - \sum_{m\in\mathcal{H}} f_{mi}(\ell) = r(\ell), \quad (\forall \ell\in\mathcal{L}, i = s(\ell)) \quad (13)$$

$$\sum_{j\in\mathcal{H}} f_{ij}(\ell) - \sum_{m\in\mathcal{H}} f_{mi}(\ell) = 0, \quad (\forall \ell\in\mathcal{L}, i\in\mathcal{N}, i\neq s(\ell), d(\ell)) \quad (14)$$

$$\sum_{j\in\mathcal{H}} f_{ij}(\ell) - \sum_{m\in\mathcal{H}} f_{mi}(\ell) = -r(\ell), \quad (\forall \ell\in\mathcal{L}, i = d(\ell)) \quad (15)$$

$$\sum_{\ell\in\mathcal{L}} f_{ij}(\ell) \leq \sum_{k\in\mathcal{K}_i\cap\mathcal{K}_j} C_{ij}^k, \quad (i, j\in\mathcal{H}, s(\ell)\neq j, d(\ell)\neq i) \quad (16)$$

$$f_{ij}(\ell) \geq 0, \forall i, j\in\mathcal{H} \quad (17)$$

$$\sum_{j\in\varphi_i^k} w_{ij}^k \leq 1, \forall k\in\mathcal{K} \quad (18)$$

$$z_{ij} \geq c(\ell) \quad (\forall \ell\in\mathcal{L}, i, j\in\mathcal{H}, s(\ell)\neq j, d(\ell) = i) \quad (19)$$

Here, σ_{ij} is the price for per unit of the allocated radio bandwidth to flow $f_{ij}(\ell)$. The flow conservation principles Equation (13), (14), and (15) specify the routing constraints. Equation (16) indicates that the total traffic routed through link $e(i, j)$ cannot exceed the total transport capacity of $e(i, j)$. z_{ij} is the storage of mobile users. Constraints (17) and (18) provide the domain constraints of variables $f_{ij}(\ell)$ and w_{ij}^k. Constraint (19) means that the storage of the accessed user z_{ij} should satisfy the demand $c(\ell)$.

To solve the multi-commodity flow (MCF) problem efficiently, we divide this problem into two parts: First, find the accessing user which can satisfy the storage constraint (including the greedy and optimal strategies); Second, find the optimal routing from the origin server to the selected accessing user. In the following parts, we give the performance analysis of these two algorithms: the Optimal-MCF algorithm (O-MCF) and the Greedy-MCF algorithm (G-MCF).

6 PERFORMANCE ANALYSIS

6.1 Parameter Settings

We simulate a 100×100 S-VDN scenario with 25 nodes using the GT-ITM tool. Each mobile user connects other users at a probability of 0.2. The storages and the link bandwidths are uniformly distributed between 50 and 100. To characterize the dynamic variation of live videos, we let the appearance of live videos conform to a Poisson process with the arrival rate parameter λ. Thus, the inter-arrival times of live videos are exponentially distributed at a mean of $1/\lambda=20$. We also assume that the duration time of the live video is exponentially distributed with parameter $\mu=500$. For each live video, the number of mobile users is set to a positive integer according to a uniform distribution between 6 and 10. The max coverage distance is set to $D_u=25$. The storage and bandwidth demands of mobile users are both set to real numbers uniformly distributed between 0 and 10.

6.2 Experimental Results

In our analysis, we have compared O-MCF with G-MCF on the metrics of acceptance ratio, revenue, cost and bandwidth utilization.

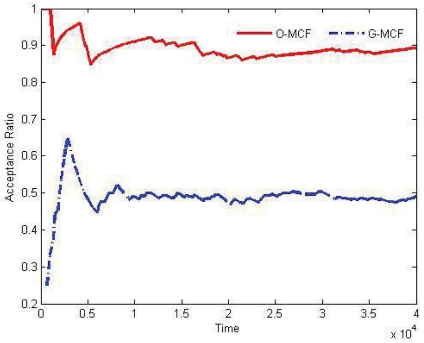


Fig. 5. Acceptance ratio vs. time

Fig. 5 depicts the acceptance ratio of live video requests for both of the two algorithms. Initially, the acceptance ratio is high because the storage and bandwidth resource is sufficient. Then, the ratio decreases along with the allocation of resources and increases along with the release of resources. Finally, the ratio decreases and stabilizes. O-MCF exhibits better acceptance ratio than other algorithms in the long run because O-MCF coordinates the access node selection with the routing constraints, which enables MCF to find appropriate routings easily. However, G-MCF only consider the access node selection according to the storage constraint.

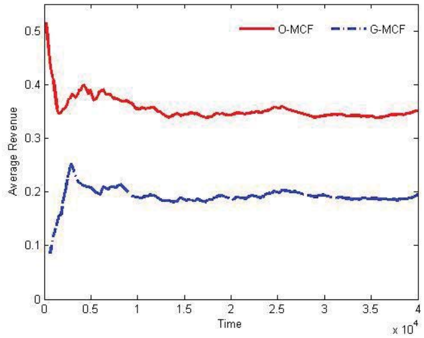


Fig. 6. Average revenue vs. time

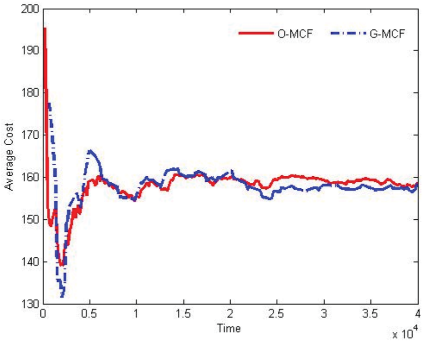


Fig. 7. Average cost vs. time

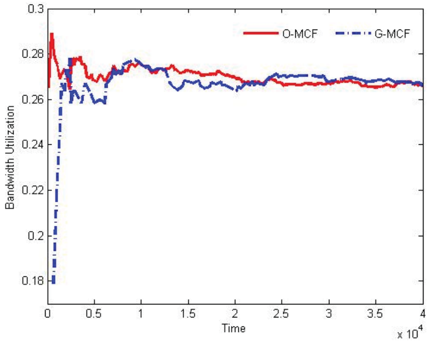


Fig. 8. Bandwidth utilization vs. time

Figs. 6 show that O-MCF significantly outperforms G-MCF in terms of the performance of average revenue as time goes on. Having more revenues along with increased acceptance ratio indicates that the proposed O-MCF algorithm actually deliver the video content that gains more revenues, instead of delivering a small video merely to improve the acceptance ratio (Fig. 5).

The results in Fig.7 and Fig.8 show that both of the two algorithms have the approximate rental cost and bandwidth utilization. Since the higher acceptance ratio leads to higher allocated bandwidth and storage, O-MCF sometimes has more cost and lower bandwidth utilization than G-MCF.

7 CONCLUSION

VDNs are used to overcome the bandwidth shortage of the Internet by deploying popular contents on proximal edge servers. However, the specific characteristics of mobile Internet, such as the mobility of mobile users and diversity of live videos, have brought huge challenges to the design of content delivery mechanisms. To address this issue, we propose a novel framework which is based on SDN and D2D technique. Then, we formulate this optimization problem as a Multi-commodity Flow (MCF) problem. Simulation experiments show that our proposed algorithms are efficient in terms of both performance and rental cost.

Acknowledgment

This work is supported by National Natural Science Foundation of P. R. China under Grant Nos. 61402270, 61373149, 61572299, 61602285, 61602284, Shandong Provincial Natural Science Foundation of China under Grant Nos. ZR2014FQ009, ZR2015FQ009, the Open Research Fund from Shandong Provincial Key Laboratory of Computer Network, Grant No. SDKLCN- 2016-02.

REFERENCES

[1] “Cisco Visual Networking Index: Forecast and Methodology, 2015-2020”, White paper, Jun. 01, 2016.

[2] Twitch is 4th in peak us internet traffic. http://blog.twitch.tv/2014/02/twitch-community-4th-in-peak-us-internet-traffic/.

[3] M. K. Mukerjee, D. Naylor, J. C. Jiang, D. Su Han, S. Seshan, H. Zhang. Practical, Real-time Centralized Control for CDN-based Live Video Delivery. In Proc. of ACM SIGCOMM, 2015, pp. 311-324.

[4] B. M. Maggs, R. K. Sitaraman. Algorithmic Nuggets in Content Delivery. ACM SIGCOMM Computer Communication Review, vol. 45, no.3, pp. 52-66, 2015.

[5] Q. Yan, F. R. Yu, Q. Gong, and J. Li, Software-defined networking (SDN) and distributed denial of service (DDoS) attacks in cloud computing environments: A survey, some research issues, and challenges, IEEE Commun. Survey Tuts., vol. 18, no. 1, pp. 602–622, 1st Quart. 2016.

[6] L. Cui, F. R. Yu, and Q. Yan, When big data meets software-Defined Networking (SDN): SDN for big data and big data for SDN, IEEE Netw., vol. 30, no. 1, pp. 58–65, Jan. 2016.

[7] Y. Cai, F. R. Yu, C. Liang, B. Sun, and Q. Yan, Software defined device-to-device (D2D) communications in virtual wireless networks with imperfect network state information (NSI), IEEE Trans. Veh. Tech.,vol. 65, no. 9, pp. 7349–7360, Sep. 2016.

[8] K. Wang, H. Y. Li, F. R. Yu, and W. C. Wei, Virtual Resource Allocation in Software-Defined Information-Centric Cellular Networks With Device-to-Device Communications and Imperfect CSI, IEEE Transactions on Vehicular Technology, Vol. 65, No. 12, pp.10011-10021, Dec. 2016

[9] E. W. Zegura, M. H. Ammar, Z. Fei, et al. Application-layer anycasting: a server selection architecture and use in a replicated web service. IEEE/ACM Transactions on Networking, vol. 8, no.4, pp. 455-466, 2000.

[10] P. Wendell, J. Jiang, M. Freedman, et al. Decentralized Server Selection for Cloud Services. In Proc. of ACM SIGCOMM, 2010.

[12] S. Srinivasan, J.W.J. Lee, D. Batni, and H. Schulzrinne. Active-CDN: Cloud Computing Meets Content Delivery Networks. Technical report, Columbia Univ., 2012.

[13] G. Rodolakis, S. Siachalou, and L. Georgiadis. Replicated Server Placement with QoS Constraints. IEEE Trans. Parallel and Distributed Systems, vol. 17, no.10, pp. 1151-1162, 2006.

[14] X. Tang and J. Xu. QoS-Aware Replica Placement for Content Distribution. IEEE Trans. Parallel and Distributed Systems, vol. 16, no.10, pp. 921-932, 2005.

[15] M. B. Tariq, R. Jain, and T. Kawahara. Mobility aware server selection for mobile streaming multimedia content distribution networks. In Proc. of International Web Workshop, 2003.

[16] Q. Xu, J. Huang, Z. Wang, et al. Cellular Data Network Infrastructure Characterization and Implication on Mobile Content Placement. In Proc. of ACM SIGMETRICS, 2011.

[17] F. Chen, K. Guo, J. Lin, and T. L. Porta. Intra-Cloud Lightning: Building CDNs in the Cloud. In Proc. of IEEE INFOCOM, 2012, pp. 433-441.

[18] C. Papagianni, A. Leivadeas, and S. Papavassiliou. A Cloud-Oriented Content Delivery Network Paradigm: Modeling and Assessment. IEEE Transactions on dependable and secure computing, vol. 10, no.5, pp. 287-300, 2013.

[19] D. Niu, H. Xu, B. Li, and S. Zhao. Quality-Assured Cloud Bandwidth Auto-Scaling for Video-on-Demand Applications. In Proc. of IEEE INFOCOM, 2012, pp. 460-468.

[20] M. L. Hu, J. Luo, Y. Wang, and B. Veeravalli. Practical Resource Provisioning and Caching with Dynamic Resilience for Cloud-Based Content Distribution Networks. IEEE Transactions on Parallel and Distributed Systems, vol. 25, no. 8, pp. 2169-2179, 2014.

[21] F. F. Chen, R. K. Sitaraman, M. Torres. End-User Mapping: Next Generation Request Routing for Content Delivery. In Proc. of ACM SIGCOMM, 2015, pp.167-181.