VIDEO REPLICATION AND ACCESS OVER FOG-BASED ARCHITECTURE

by

ZHANGYU CHANG

A Survey Paper Submitted to
The Hong Kong University of Science and Technology
in Partial Fulfillment of the Requirements for
Ph.D. Qualifying Examination (PQE)
in Computer Science and Engineering

May 2018, Hong Kong

TABLE OF CONTENTS

Title Page **i**

Table of Contents **ii**

List of Figures **iv**

List of Tables **v**

Abstract **vi**

Chapter 1 Introduction **1**

Chapter 2 System Architecture and Comparison **6**

2.1 System Architecture of Fog-based Video Network 6

2.2 Comparison 7

2.2.1 Cloud-based CDN 9

2.2.2 Peer-to-Peer 9

Chapter 3 Replication Schemes **11**

3.1 Uncoordinated Schemes 11

3.1.1 Variations of LRU 11

3.1.2 Score-Based Schemes 12

3.2 Coordinated Schemes 12

3.2.1 Popularity-Based Schemes 13

3.2.2 Division of Storage 14

Chapter 4 Video Access Schemes **15**

4.1 Wired Users 15

4.1.1 Clustering Methods 17

4.1.2 Game Theory Approaches 18

4.2 Wireless Users 18
4.2.1 Approximation Algorithm 19
4.2.2 Mathematical Programming 21

Chapter 5 Case Studies 22
5.1 Youku: CDN Based on Smart Routers 22
5.1.1 Combining Replication with Recommendation 23
5.1.2 Centralized User Access Coordination 23
5.2 Thunder Crystal: Crowdsourcing Content Distribution 25
5.2.1 Replication Based on Popularity 25
5.2.2 Random User Access 26

Chapter 6 Conclusion and Future Directions 28

References 30

LIST OF FIGURES

1.1 Hierarchical view of cloud, fog and user devices. 2

2.1 Fog-based video distribution platform. 8

4.1 A fog-based distributing architecture in wired network. 16

4.2 A wireless fog architecture with store-capable base stations. 20

5.1 Architecture of the peer CDN system of Youku. 24

5.2 System architecture of Thunder Crystal. 27

LIST OF TABLES

2.1 Different paradigms for video distribution. 7

3.1 Different uncoordinated replication schemes. 12

3.2 Different coordinated replication schemes. 13

4.1 Different wired user access schemes. 17

4.2 Different wireless user access schemes. 19

VIDEO REPLICATION AND ACCESS OVER FOG-BASED ARCHITECTURE

by

ZHANGYU CHANG

Department of Computer Science and Engineering

The Hong Kong University of Science and Technology

ABSTRACT

Efficient video content replication and distribution is essential for video service providers to support a huge and increasing demand of users and ensure the quality of experience (QoE) with reduced operation cost. As the nowadays ubiquitous and close-to-user fog devices (Wi-Fi routers, set-top boxes, small cell base stations, etc.) have better computation, bandwidth and storage capacities, fog devices can collaboratively serve the local users and provide a new paradigm for video distribution. However, coordinately managing millions of geo-distributed fog devices required new schemes that are different from both the traditional content delivery network (CDN) or the peer-to-peer (P2P) schemes.

This survey overviews the recent advancement in fog computing to support video service. We discuss the recent work on the video popularity characteristics, the replication strategies and the access schemes for fog-based video service platform. We first overview the video popularity characteristics and discuss the fundamental challenges to design an effective fog-based video distribution architecture. We then examine the recent advances on two major areas: video replication and access. Regarding the replication problem, we present both

uncoordinated and coordinated replication schemes for a large pool of geo-distributed fog devices. Regarding the video access problem, we discuss how the users can reach their desirable video contents for both wired and wireless networks. We then study 2 real implemented cases (*Thunder Crystal* and *Youku Smart-router*) of existing large-scale fog-based content distribution platform and discuss several future directions.

CHAPTER 1

INTRODUCTION

Video contributes to most of the internet traffic and the weight is continuously increasing. It is estimated that video traffic will be 82 percent of all consumer traffic by 2021, up from 73 percent in 2016 [10]. Traditionally, Video on Demand (VoD) service is heavily relied on the cloud side, especially on Content Distribution Networks (CDNs). The video content providers have to purchase more resource from CDN operators to meet the exponentially growing user demand, which is not cost-effective. Due to the huge network resource used for video service, any new paradigm that can offload such demand would have a huge impact.

As many nowadays so-called lightweight devices that connect to the Internet are equipped with decent computing, bandwidth and storage capacities (such as routers, Wi-Fi access point, set-top boxes, small-cell base stations, etc.), such devices are able to provide many types of service that can only be offered by the dedicated servers in the past. Therefore, a new computing paradigm based on these devices emerges with various names, such as fog computing, edge computing [31] and crowdsourcing computing [9]. Fog devices also have many nicknames, such as edge devices [25] and micro/nano data centers [46]. Despite the various names, the key philosophy is to use lightweight (but already powerful enough) and close-to-user devices to serve the users with lower cost and better responsiveness. In this work, we use fog computing to represent all such related terms.

We show the hierarchical view of cloud, fog and user devices in Figure 1.1. Fog devices bring 2 features to the video distribution service. First, fog devices can be deployed ubiquitously, so the load of the cloud-side servers can be reduced. Also, fog devices are deployed closer to the users, so the link distance to transmit the video to the user is small. By serving the users with the fog devices, the content providers can reduce the cost of both servers and network traffic.

Figure 1.1: Hierarchical view of cloud, fog and user devices.

Any distribution paradigm has to consider and utilize the video popularity characteristics. To understand how to effectively deliver video to the users, we study video popularity characteristics, which include video popularity, freshness and daily request pattern.

- *Popularity*: Various work [21] has confirmed that video popularity is very skewed for both the video made by both *professionally generated content* (PGC) and *user generated content* (UGC). For the famous user generated content platform such as YouTube and Daum (a popular site in Korea), it is shown that both of their video popularity patterns follow the Zipf distribution with exponents between 1.5 (Daum) and 2.5 (YouTube) for the popular videos. Specifically, by storing only 10% of long-term popular videos, a cache can serve 80% of requests [5]. Similar results are confirmed by many works that measure the user preference [9, 21, 31, 33]. Though both PGC and UGC have the long tail problem, the videos in the long tail are seldom stored in the fog devices.

- *Freshness*: As the video popularity decays very quickly with time, the freshness of the video is also a very important factor when deciding which content to replicate. Statistics from various video service platforms [5, 31, 33] indicates that the popularity of hot videos decays very quickly. For PGC, nearly 90% of the most popular videos (i.e., the top 10 percentile) are new videos each day. Therefore, for any content distribution system, it has to push new contents to the fog devices every day. For UGC, it is difficult to predict the popularity of a new content.

- *Daily Pattern*: The end user demands pattern follows a daily cycle. The demand pattern has 2 peaks at around 2 P.M. and 10 P.M. every day. The traffic reaches its lowest at around 5 A.M. daily. It is preferred to push the new contents when the user demand reaches its daily minimal.

Though fog-based content replication and delivery is promising for video service, it faces challenges that have not been addressed by the previous work on cloud-based CDN or Peer-to-Peer (P2P) paradigm. In order to efficiently utilize fog-based distribution platform, the fog operators have to carefully design new collaborative strategies and address the challenges raised by the video popularity characteristics. To meet the dynamic changes on user demand, effective video distribution system has to timely update its video replication and access decisions.

Furthermore, due to the intrinsic nature of fog-based network, the replication schemes must show the following features:

- *Distributive*: Fog devices are huge in number and have to collaboratively serve the users.
- *Geography-aware*: As fog devices are generally close to the users, it is important to utilize such feature so that the fog device serves its nearby users.
- *Popularity-aware*: To effectively use the fog devices and reduce the load of the cloud-side CDN, the popular contents have to be pushed into the fog.
- *Lightweight implementation*: Fog devices, though their computation power has improved in the recent years, still cannot compare to the dedicated servers in terms of computation power. The schemes for fog devices have to meet such requirement.

Due to these issues, existing replication schemes that perform well in the traditional cloud-based CDN or P2P paradigm may not be directly applicable to fog-based architecture.

In this work, we review the work on fog-based video replication and distribution. Specifically, we focus on the work to address the video replication and access problem. These challenges can be categorized into 2 major areas: video *replication* and *access* problems. In other words, what content items should be replicated to which fog device? How to organize the distributed fog devices?

- *Replication*: Fog devices are huge in number, but each fog device has only limited storage size. Video storage decision for each fog device has great impact on the performance of the fog platform. Effective video replication relies crucially on the knowledge of content popularity. Fresh PGC such as news, music or TV series is produced on a regular basis. One of the characteristics of such content is that it is ephemeral. I.e., it is highly demanded for a certain duration and then the demand fades. Therefore, the fog operator has to decide not only what to store in the fog device, but also what to push and when to push the new content.

Current replication schemes for fog devices fall into 2 categories: *uncoordinated* and *coordinated* replication schemes. For uncoordinated schemes, each fog device only makes its own decision based on the information it collects from previous history to serve the users. For coordinated schemes, fog devices make the decision based on the global popularity information.

- *Video Access*: An individual fog device cannot have all the videos due to its limited storage. Meanwhile, many fog devices may have the same contents across different region. Therefore, a user has to choose among many fog devices for a video. A user's choice on which fog device to access the video not only affects the experience of its own, but also affects other users share the same resource (e.g., same fog device and links).

  For *wired* users, the management scheme has to coordinate with millions of devices. The key problem is to deal with the huge problem size. For *wireless* users, as the coverage of the base stations may overlap, it is essential to decide how to choose the base stations to access the video for each user.

The remainder of this paper is organized as follows. In Chapter 2 we overview the system architecture of the fog-based video network and compare it with the traditional CDN and P2P paradigms. We discuss the replication problem in in Chapter 3, and the video access problem in Chapter 4. We then study two commercially implemented cases in Chapter 5. We finally conclude and discuss the open research directions in Chapter 6.

CHAPTER 2

SYSTEM ARCHITECTURE AND COMPARISON

We present the system architecture of the fog-based video network (Section 2.1) and compare it with the traditional paradigms (Section 2.2).

2.1 System Architecture of Fog-based Video Network

In Fig. 2.1, we show the basic system architecture of a fog-based video distribution platform. We can decompose it into 3 layers:

- The *Cloud Layer* (Level 1): The video content providers still need some cloud-based CDN servers. However, such servers seldom serve the user directly. Instead, the major role of these servers is to transmit new videos to the fog devices so as to update the replication in their storage. Therefore, the content provider does not need too many cloud CDN servers and the cost on cloud CDN is reduced.

- The *Fog Layer* (Level 2): Fog devices (e.g., routers, switches, set-top boxes, base stations, etc.) are the backbone of the platform and act as an intermediary between the cloud and user devices. Their computing, bandwidth and storage capacities are not as powerful as the dedicated servers in the cloud side, but they have a huge number and are close to the users, usually in their homes. The fog devices can support both wired and wireless users. Note that the owner of the fog device and the operator of the fog device may not be the same. A user can own the fog device and put it in its home, and let a fog operator to use it in exchange for better quality of video service or monetary return.

Table 2.1: Different paradigms for video distribution.

	Cloud CDN	P2P	Fog
Data Storage	Centralized	Distributed	Distributed
System Control	Centralized	Uncoordinated	Coordinated
QoS	Yes	No	Yes
Capital Cost	High	Low	Low
Scalability	Low	High	High
ISP Friendly	Yes	No	Yes
User Contribution	No	Required	Desirable

In a *wired fog*, a fog device acts as a mini server to serve the neighbor user requests. In a *wireless fog*, a small-size base station (i.e., a small cell or femtocell base station) with some storage and processing power can serve the users within its signal coverage.

- The *User Layer* (Level 3): End user devices usually get served by fog devices. Upon a user request, the fog devices near the user will directly serve such request if they have the content. Occasionally, if all the nearby fog devices do not have the desired video content or do not have the bandwidth, a user may also get served by the cloud CDN server.

In general, the cloud layer pushes the new contents to the fog layer. In fog layer, fog devices not only serve the user requests, but also deliver the new contents to each other. Users usually get the video from a neighbor fog device, but may get served by the cloud if no fog device has the desired video.

2.2 Comparison

Fog-based video distribution architecture has very distinctive features compared with the traditional cloud-based CDN and the peer-to-peer network. In general, fog-based distribution shares the merits of both the cloud-based CDN and P2P model but avoids the drawbacks of them. We compare the fog based architecture with CDN (Section 2.2.1) and P2P (Section 2.2.2). Table 2.1 compares the major difference of the paradigms used for video distribution.

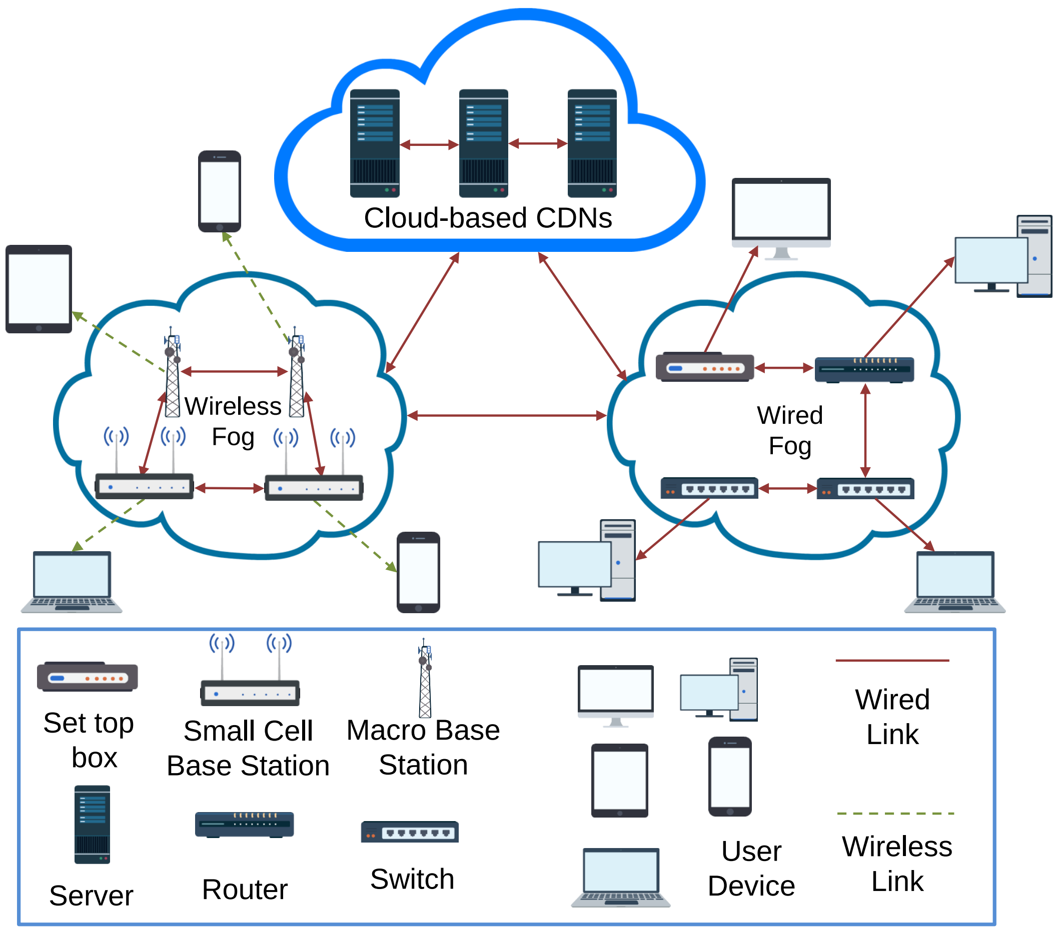

Figure 2.1: Fog-based video distribution platform.

2.2.1 Cloud-based CDN

Compared with cloud-based CDN [6, 7, 34], fog-based distribution uses fog device as the backbone, which leads to lower cost and better responsiveness. The reduction of the cost comes from various factors.

- Better proximity: As the fog devices are distributed close to the users, the bandwidth cost between the cloud and the user can be saved. The bandwidth cost to update the content in the fog storage is also reduced as the cloud layer only needs to push the new contents to a fraction of the fog devices (e.g., 10% of the fog devices within each ISP domain [33]), and these fog devices will further distribute the new contents to the other fog devices.

- Reduced cost: As the cost of the storage and processing hardware has declined constantly, nowadays the major cost related to cloud-based CDN has been shifted to real estate, power, cooling and human resource, which do not exist on fog devices [22, 25]. In this sense, fog computing is a green computing paradigm.

- User contribution: As the close-to-user fog devices can reduce the transmission delay and enhance the quality of experience, users are willing to buy Wi-Fi routers that support fog computing even without subsidy from the content providers. Some content also providers pay its users to install particular Wi-Fi routers to expand their fog network [33].

2.2.2 Peer-to-Peer

Compared with P2P approach [28, 42, 51], the content providers can effectively control and coordinate fog devices. Therefore, fog computing can overcome some inherent limitations of the P2P paradigm and has some distinctive merits.

- Service guarantees: By controlling the fog devices, the fog operator can build stable network connections and reduce the peer churns of fog devices. Each device has a higher standby probability and has dedicated resources to support the video service. Even for unpopular content, we can store them in some fog devices where the user may not request it at the moment.

- Coordinated topology: The topology of the fog network is more manageable. The content provider can set the rules on the video distribution based on the ISP information. Consequently, this management can reduce the inter-ISP traffic and the traffic distance between the video source and the user.

- Free-riding prevention: The fog device can still store and distribute the contents even the house owner is not interested in them. As the users have to rely on the fog devices for other functions (e.g., Wi-Fi, television), they have less incentive to turn it off.

Note that in some of the crowdsourcing computing schemes [9], the video content providers pay the user to support the platform. In such case, personal computers (desktops and laptops) can also be in the fog layer. This is clearly different from the P2P approach since the content providers get full control of the user resources after the payment. Therefore, the drawbacks of the P2P can be avoided in such case.

CHAPTER 3

REPLICATION SCHEMES

Replication scheme for fog must be distributive, responsive and easy to implement. For *uncoordinated replication* schemes, each fog device only makes its own decision independently based on the information it collects from previous service history. Namely, it only knows what the user has requested from itself and decides what to store and what to replace when a new user request comes. Such schemes are simple and distributive, but lack global popularity information. On the other hand, in *coordinated replication* schemes, a central server will offer global information to each fog device in some way, or even push the contents directly to it. However, such decision may ignore the local preference of the video popularity. We present uncoordinated replication (Section 3.1) and coordinated replication (Section 3.2).

3.1 Uncoordinated Schemes

We present variations of Least Recently Used (LRU) scheme (Section 3.1.1) and score-based schemes (Section 3.1.2) that propose new benchmarks to make the replication decisions. Table 3.1 compares different uncoordinated replication schemes.

3.1.1 Variations of LRU

Least Recently Used (LRU) and Least Frequently Used (LFU) are the most commonly used replication schemes. Some work that majorly addresses the video access problem use them as the default replication schemes.

There are some schemes that modify LRU [15]. In *q-LRU*, the fog device stores the new video content with a probability of q upon a request. In *k-LRU*, the storage is divided into k hierarchical parts. Each part demotes its least recent used content to a lower level.

Table 3.1: Different uncoordinated replication schemes.

Scheme	Objective	Parameter to optimize	Methodology	Comment
LRU-based [15]	Maximize hit probability	Video stored in the device	Poisson Approximation	Assume Zipfs Law
iProxy [14, 43]	Improve hit probability	Video stored in the device	Heuristics	New coding scheme
Age-based Threshold [26]	Maximize hit probability	Video lifetime in the device	Poisson Approximation	Assume Zipfs Law

Therefore, only the content in the lowest level can be removed from the storage. Both *q-LRU* and *k-LRU* reduce the frequency of the changes in the storage, and performs better than LRU if the video popularity distribution follows the Zipf's law.

3.1.2 Score-Based Schemes

The work in [26] proposes a score-based replication scheme. To decide which content to store, it considers an Age-Based Threshold (ABT). Given the time t a content has been stored in the cache and a function N, a content has to satisfy $N(t) > N_{min}$ stay in the cache, where the fog operator defines N and N_{min} to fit the popularity and user access pattern. This work shows that the scheme is close to the optimal solution if the user request rate follows the Poisson distribution.

For Information Centric Network [14, 43], content providers can use the Information-Bound Referencing to index the same video with different resolutions. To achieve dynamic video encoding, this work proposes a new coding scheme (with frequency domain data) for multi-resolution video with less storage. This work uses a score-based replication algorithm (LFU-based IBR-score). It differs from LFU in that each access has a different weight, and the latest demand has a higher score.

3.2 Coordinated Schemes

We present popularity-based schemes (Section 3.2.1) which use the global popularity information for replication decision, and schemes that divide the storage of the fog device

Table 3.2: Different coordinated replication schemes.

Global Popularity [8, 52, 53]	Reduce server load	Video stored in the device	Mathematical modelling	Not consider bitrate/device capacity
Deficit Bandwidth [48]	Reduce server load	Video stored in the device	Mathematical modelling	Consider device capacity
Last-mile Implementation [23]	Reduce traffic cost	Probability to store a video	Primal-dual approach	Comprehensive model
Social Video Index [47]	Improve hit probability	Video indices	Heuristics	Based on measurement
Division of Storage 1 [31]	Maximize hit probability	Storage Division	Heuristics	Support Wired/Wireless users
Division of Storage 2 [19]	Maximize social welfare	Storage Division	Supermodular game	Support social video

(Section 3.2.2) to support different types of videos. Table 3.2 compares different coordinated replication schemes.

3.2.1 Popularity-Based Schemes

With global information, some researches show that they can do better that LFU/LRU. Some work [8] states that the global population of the video content shall be proportional to the video popularity. The work in [53] pushes the contents to the fog devices, and the number of replicas for each video is based on the global popularity. The work in [52] shows that replication shall be proportional to the demand to achieve the optimal performance, and validate that the proposed schemes outperform LRU and LFU through simulation.

To effectively carry out the decision, the work in [23] proposes a solution based on set-top boxes for last-mile CDN. It clusters the fog devices based on their ISPs and replicates the contents based on global information. In each ISP, it deploys a tracker which collects the popularity information. Tracker gives replication probability p and broadcast this parameter to all the fog devices. The fog devices distributively compute the replication decisions based on that p. Namely, each fog device independently replicates the video with a probability p. Therefore, the global replication number reaches the expected value.

However, the work in [48] shows that proportional replication is not optimal in P2P-VoD as fog devices may have different uploading capacities. Deficit bandwidth performs better than proportional replication. If popular contents do not have enough bandwidth to support the users, they shall have more replicas. This replication strategy can be used in both uncoordinated and coordinated replication schemes.

The work in [47] proposes another benchmark for replication instead of global popularity. By using cloud-fog architecture to support social video, it defines 3 indices (geographic influence/content propagation/social influence) to reflect the geographic, time and popularity feature of a video. However, it only uses the geographic closest fog device to access the video.

The work in [37] gives a mathematical approach to optimize the replication decision. Given the video popularity, it uses the primal-dual method to calculate the *Time to Live* for every video. The time-to-live feature can be a tag for the video content. Each fog device stores and replaces the video based its time to live.

3.2.2 Division of Storage

The work in [31] analyzes the user request temporal and spatial features, and proposes a replication strategy to support both wired and wireless users. For the storage of each fog device, it shall be divided into 2 parts. One is to serve the single-location user based on local popularity. The other is for mobile user based on global popularity.

The work in [19] suggests that fog devices may use extra storage to propagate the user generated social videos. It shows that the users have incentive to upload video for their friends through supermodular game approach. The major goal of the problem formulation is to maximize the total social welfare given the upload traffic constraints for each fog device.

The work in [50] proposes Local Hardware Awareness (LHA), a special API, to find the nearby storage devices. Each fog device divides its storage into 2 parts for virtual machine (VM) and storage. A fog device can quickly load the different VM or contents from the nearby storage to respond to the changes of user demand.

CHAPTER 4

VIDEO ACCESS SCHEMES

Video access decision is critical to effectively utilize fog network resources and avoid network congestion. With a large number of fog devices, it is possible that there are multiple replications for the same video content in several fog devices, and we have to decide from which device a user shall get the video content. As the number of both the videos and fog devices are huge, the size of the problem to optimize the whole network is huge. Furthermore, the video access problem can be regarded as *Facility Location Problem*, which is NP-complete. To effectively solve the problem, it is important to reduce the problem size and find the approximation algorithms. We present video access schemes for wired users (Section 4.1) and for wireless users (Section 4.2).

4.1 Wired Users

We show a fog-based distributing architecture in wired network in Figure 4.1. For *wired* users, the management scheme has to coordinate with millions of devices. Such problem is also NP-hard in nature, so it is unlikely to find the optimal solution in real time. To reduce the problem size, a common approach is *divide and conquer*. The fog operator usually divides the whole system into several regions and solves the small size problems within each region. The key problem is how to cluster the users and divide them into regions. Table 4.1 compares different wired user access schemes.

Some work uses simple methods for small scale problem. The scheme in [40] is a simple VoD sharing mechanism that a user will ask neighbor fog devices for help when requesting a video. The focus of this work is to design a proof-of-concept system. The authors implement a prototype system and test it on 4 desktop computers with 30 videos. The major metrics of the experiments are delay and throughput. The work in [24] proposes a brutal force method

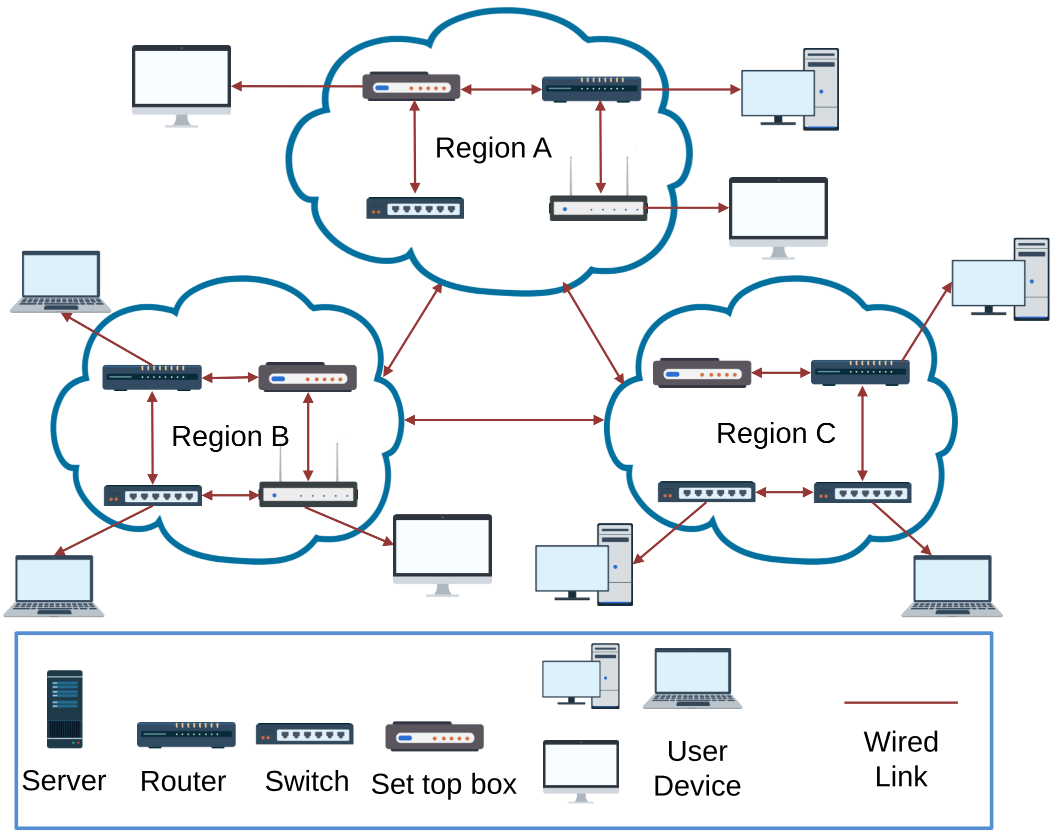

Figure 4.1: A fog-based distributing architecture in wired network.

Table 4.1: Different wired user access schemes.

Scheme	Objective	Parameter to optimize	Methodology	Comment
Cluster fog devices 1 [20, 44]	Reduce server load	How to partition fog devices	Sampling and greedy algorithm	Based on measurement
Cluster fog devices 2 [32]	Reduce server load	Traffic between clusters	Linear programing and Heuristics	2 schemes can be combined
Game theory approach 1 [27]	Total revenue	Price to use a fog device	Stackelberg Game	Fog owner and CP can cooperate
Game theory approach 2 [48]	Maximize social welfare	From which friend to get video	Supermodular game	Assume friends share videos

to decide whether to access video from the nearest local fog device or from the cloud server. It shows that the replication method of LRU, GLRU, LFU and GLFU are better than random method.

For large scale problem, recent work usually uses clustering methods (Section 4.1.1) and game theory approaches (Section 4.1.2).

4.1.1 Clustering Methods

Some work [44] tries to use clustering method to reduce the problem complexity of video access. Besides measurement study, this work proposes to partition the fog devices according to geographic location. Inside each region (usually within an ISP domain), the content providers can put a coordinate server to manage the fog devices inside the region. This work shows that within the region, the popularity characteristics are usually very similar. It further proposes auction-based method to let the fog device owner get payed from the end user. The work proposed in [20] uses a similar clustering method for video access but use LRU for replication.

The work proposed in [32] addresses the problem of cooperation between regions. Instead of solving the minimum-cost-maximum-flow (MCMF) problem, this work proposes that, after clustering the users into regions, the fog operator can compute the Jaccard similarity of the user demand between regions. Cooperation between the regions of high similarity can increase the hit ratio and reduce the delay. The algorithmic complexity is lower than LP-based solution.

4.1.2 Game Theory Approaches

The work in [27] proposes a game theory approach to show that the content provider (CP) and the fog owners can reach equilibrium and mutual benefit. It uses Stackelberg Game (with equilibrium) to improve the total revenue. With such game, the content providers can reduce the cost and the fog device owner can maximize their revenue. However, it just uses most popular first (MPF) for replication and each user is only access the video from one device (usually the closest one).

The work in [48] proposes that the fog devices shall use extra storage to propagate the social video. The collaboration of fog devices is based on the social network. It shows that the users have incentive to upload video for friends with supermodular game approach. With this game approach, the network can maximize social welfare even if the upload traffic constraints exist.

4.2 Wireless Users

We show a wireless fog architecture with store-capable base stations in Figure 4.2. Base stations and servers can transmit video to each other through wired links, and a mobile user device can get video from a based station through a wireless link. For *wireless* users, the coverage of the base stations may overlap. For example, user devices A, B and C are covered by multiple base stations. The major problem to solve is to choose the proper base station to serve the users when multiple base stations cover the same region. Table 4.2 compares different wireless user access schemes.

For a wireless network, a base station with video storage receives requests from a small population of users. Hence, the number of requests per unit time is also very small. Meanwhile, as the coverage areas of base stations or Wi-Fi access points are overlapping, the choice of the base station/access point has great impact on the efficiency of the data transmission.

Table 4.2: Different wireless user access schemes.

Scheme	Objective	Parameter to optimize	Methodology	Comment
JRC-UR [39]	Minimize server load	Replication and Access	Approximation of LBS	Approximation ratio given
BS assisted D2D [17]	Minimize server load	Replication and Access	Monte Carlo optimization	Heuristics in nature
AP deployment [3]	Minimize server load	Fog device deployment	Integer linear programming	With a greedy heuristics
An online algorithm [36]	Minimize server load	Replication and Access	Convex programming	Allow user to access many APs
FemtoCaching [41]	Minimize server load	Replication and Access	Linear Programming	Use coded content

Compared with wired fog, the problem size of a wireless network is small. Therefore, current work usually considers jointly optimize the replication and access problem together with approximation algorithms (Section 4.2.1) or mathematical programming (Section 4.2.2).

4.2.1 Approximation Algorithm

For the work related to small cell network [39], the authors use local storage at the base station and propose the joint optimization problem of video replication and access. It aims to formulate both the video replication and access together with a mathematical model and solve the linear programming problem. As the problem is NP-hard, this work tries to modify the problem as a facility location problem, and uses a related approximation algorithm to get the solution. However, how to implement the decisions from the mathematical programming is still an open question.

The work in [16] tries to solve the similar distributive caching in femtocell base station. It proposes an algorithm that clusters user demand to reduce the problem complexity.

Another work [17] considers the device-to-device model within a single-cell of cellular or Wi-Fi network. Smartphone users within the signal coverage aim to help each other for video downloading. The base station acts as a controller/scheduler to coordinate the sharing. This work shows that randomized replication is not bad. However, in the setting, it only has a small video and user pool with only 1000 videos and 500 users, which are far from the reality.

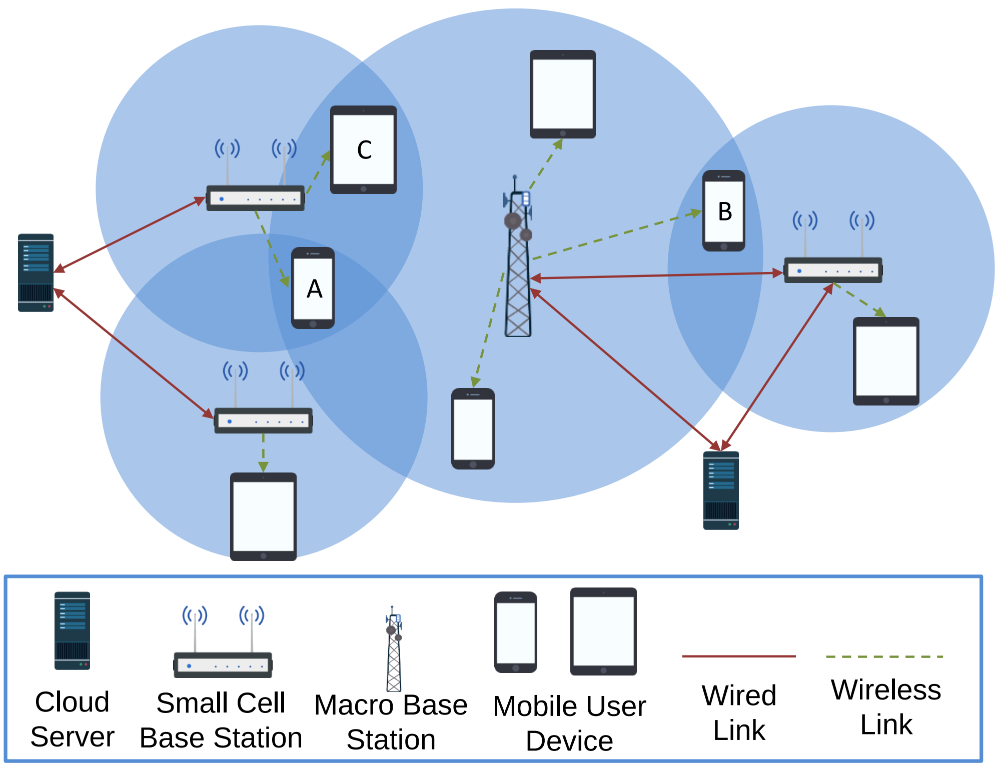

Figure 4.2: A wireless fog architecture with store-capable base stations.

The work in [3] examines the problem of choosing the location to deploy the wireless fog device (e.g. Wi-Fi access point, small cell base station). This work proposes 2 algorithms for deployment. The first one is a greedy algorithm which deploys the fog devices in the position to cover the maximum number of users. The second one is based on integer linear programming which can give the optimal solution. The simulation results show that the greedy algorithm performs quite close to the optimal solution, but has smaller algorithmic time complexity.

4.2.2 Mathematical Programming

To solve the similar problem, the work in [36] also tries to solve the content replication and request assignment problem jointly. It tries to formulate the problem as a convex programming problem by allowing a user to get the contents from different base stations with a given ratio close to the solution of the convex program. It shows that if the request number is large, the network performance goes close to the optimal. Besides, it gives an online algorithm for each fog device to implement the replication decision. The basic idea is to find which operation could improve the system most and do the most desirable operation first.

The work in [41] considers using coded content to bypass the NP-hardness of the integer linear programming. A video content can be encoded into M coded contents where a user can recover the video from any N coded contents, where M and N can be set by the fog operator. It proposes LP-based algorithm for coded contents, and shows that the network performance goes to the optimal if N is large.

Similarly, the work in [38] considers caching multi-layer video in small-cell base stations. As multiple base stations may cover the same area, it jointly considers caching and routing via integer convex programming. To solve the problem, the mathematical model can be decomposed into sub-problems, and each fog device can solve its own sub-problem independently by primal-dual approach.

CHAPTER 5

CASE STUDIES

Some major video content providers [1] and startup companies [4, 30] have used fog-based video distribution platform. In this chapter, we study 2 commercially deployed fog-based video content distribution systems: *Youku Smart-router* (Section 5.1) and *Thunder Crystal* (Section 5.2). Both of them support wired network and deploy the fog resource in users' home, but they still have some differences.

5.1 Youku: CDN Based on Smart Routers

Youku, one of the largest online video content providers in China [29], has deployed over 300,000 smart-routers in the homes of its end users, expecting that a large fraction of such fog devices can act as content delivery peer nodes. In this peer CDN system with agents (fog devices), Peer CDN control servers and CDN Infrastructure, the HTTP protocol is used to download contents from edge servers, and a private P2P protocol based on UDP, is adopted for the peers to deliver their data. A smart-router downloads the content from multiple peers in parallel when it obtains the peer list. We describe the approaches Youku uses to deal with the major challenges as follows. Youku has not yet published how they operate the fog network. The understanding of its fog comes from the measurement work to monitor the traffic of Youku's router [33].

Youku subsidized its users to deploy the smart-routers. Each smart-router has additional functions (e.g., set-top box, Wi-Fi access points) with lower price. Users with the smart-routers can also get extra benefits (reduction in membership fee) from uploading the contents with users' bandwidth. We describe the approaches Youku uses to deal with the major challenges on video replication (Section 5.1.1) and access (Section 5.1.2) as follows.

5.1.1 Combining Replication with Recommendation

Youku uses a centralized content replication strategy. Both the content push and content removal are scheduled by the centralized control. Video popularity is the most critical influential factor in the video placement. The videos that are popular in the recent month are likely to be replicated in smart-routers (e.g., over 60% of the peer-videos are among 1200 most popular videos of the recent month). Youku content replication strategy is insensitive to the local content popularity and scheduled globally.

In addition, Youku combines the replication strategy with its recommendation system. For example, the home page of Youku generally presents the newly published videos, recommended videos and popular video links, which can attract users to click. It is observed that 73% of the videos stored by our peer routers are on the home page, and TV series and cartoon have the highest possibility to be stored by smart-routers when they are recommended.

The downloaded chunk number follows a daily pattern. Specifically, the lower levels of chunk downloaded happen between 5 and 7 A.M. and 2 and 4 P.M. every day, and the peaks happen between 0 and 3 A.M. This indicates that chunk downloading is scheduled periodically (i.e., hourly) during a day. It is likely that Youku makes replication decision on a daily basis and push the contents during the time when the Internet traffic level is low.

5.1.2 Centralized User Access Coordination

We show the architecture of the peer CDN system of Youku in Figure 5.1. The coordinating system in Youku is highly centralized. Youku has 4 kinds of servers to manage the smart routers:

- Config servers: Peer routers download configuration parameters from Config servers.
- QoS monitors: Peer routers report statistics to the QoS monitor server, including the information of their partners and their operation states.
- Scheduling servers: Scheduling servers schedule the content replication according to the information monitored.

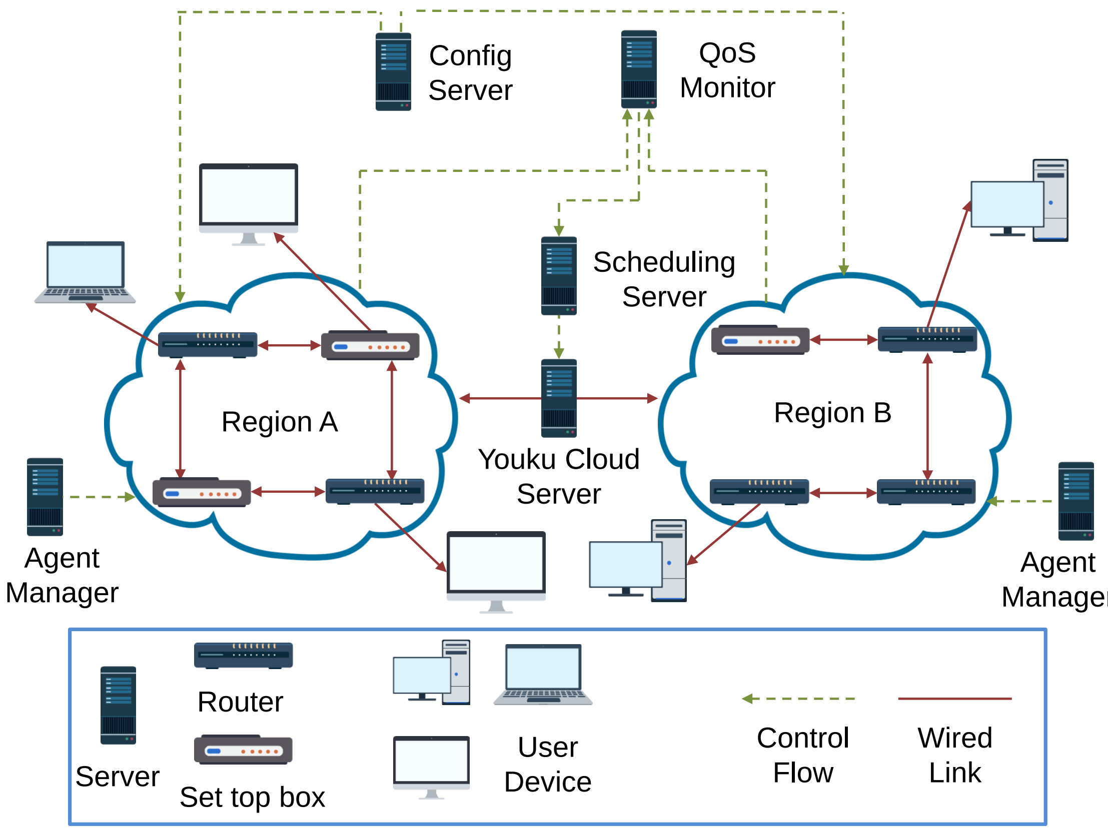

Figure 5.1: Architecture of the peer CDN system of Youku.

- Agent managers: A set of agent managers is deployed in different ISPs and locations, where each of them manages the smart-routers that are close to it.

As it seems that Youku do not care about the local content popularity, the control servers can give similar instruction on all the fog devices. The fog devices usually store the similar popular contents, so the users only need to find the nearest available fog device for the video contents. The agent managers can make sure that users mostly request the content with the same ISP domain.

5.2 Thunder Crystal: Crowdsourcing Content Distribution

Thunder (Xunlei) is a popular P2P file sharing client side software in China [12,13]. To extend its business, the users (who are called *agents*) with installed Thunder software can help some video content providers to distribute the contents for monetary return. The operator of Thunder Crystal, together with some researchers, has published some work to illustrate how Thunder Crystal works [9].

In this work [9], the Thunder Crystal network is called a *crowdsourcing system*. In this network, Thunder asks some users (agents) with surplus bandwidth and computing resources to support its content delivery. Unlike traditional P2P paradigm, Thunder has greater access to the agents' devices if the agents agree to contribute their storage (to store content) and upload bandwidth (to distribute content to end users). Each agent acts as a mini-CDN edge server, which is very close to end users. Thunder's cloud servers have the right to keep pushing content to agents, and Thunder rebates cash to its agents.

Most agents are normal Internet users who would like to offer their surplus bandwidth with very low charge. A user can have a better price by contributing their bandwidth in peak hour. With the current price, it is enough for a user to cover its electricity and network utility bills. Compared to the charging policies of the CDN companies, the cash rebated to agents is much lower, implying lower expenses for distribution service. We describe the approaches Thunder uses to deal with the major challenges on video replication (Section 5.2.1) and access (Section 5.2.2) as follows.

5.2.1 Replication Based on Popularity

In Thunder's push strategy, agent devices are not discriminated. Thunder servers push the new replicas randomly to different agent devices. Therefore, the fog network has no locality awareness, and the agents in the fog do not cooperate on contents push. Thunder servers can push 80TB traffic per day as their traffic budget. Files will be pushed out by decreasing order of popularity until the traffic quota is used up. Therefore, more popular files will be pushed with higher priority than less popular files. If the storage of an agent is full, the file with the lowest global popularity will be replaced.

To calculate the global popularity, besides the new contents, if the number of requests to download a file is N, Thunder Crystal will maintain $(0.05N)^{0.96}$ replications in the next day. Such formula is obtained from experimental trials.

5.2.2 Random User Access

In Figure 5.2, we show the system architecture of Thunder Crystal. The Thunder servers get the prediction of video popularity from a scheduler and get the new video contents from the content provider. The servers push the contents to the fog devices indiscriminately. I.e., Thunder Crystal does not have the geography awareness.

Currently, Thunder Crystal system has more than 11,000 active agent devices, which is much smaller than the scale of Youku. Despite the proximity of the agents, Thunder Crystal uses a random video access scheme that undermines the ISP friendliness. For each agent device, it serves all received requests with best efforts. For a user request, the file downloading is conducted chunk by chunk with each chunk size 2MB. The first chunk is from the cloud CDN, and the rest of the chunks will be chosen from around 200 agent devices with the desired content randomly.

A data report module is embedded into the software that installed on the agent devices. Reported information includes the number of agent devices that are working at any moment, measured instant uploading speed, event-driven messages recording the access log of content and other activities. However, as the push strategy of Thunder Crystal is indiscriminate, it is not clear how the information affect the push decision. To protect the copyright and avoid the user to falsely report its contribution, all the files on any agent's fog device are encrypted.

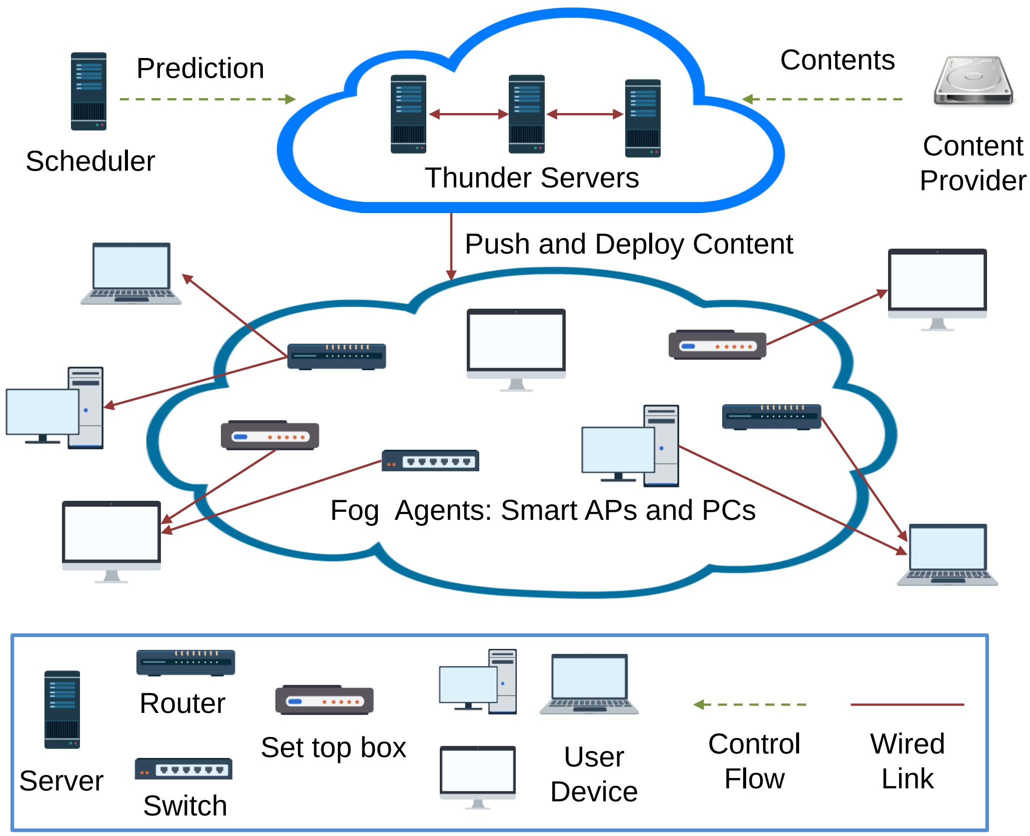

Figure 5.2: System architecture of Thunder Crystal.

CHAPTER 6

CONCLUSION AND FUTURE DIRECTIONS

Due to low operation cost and better quality-of-service, both researchers and the industry regard fog-based video content distribution as a promising solution to the skyrocketing growth of the demand of video service. Unlike the traditional cloud-based CDN or P2P paradigm, fog-based architecture has good scalability, reduced cost and guaranteed quality of service. We conduct this survey to review the recent work.

We first briefly go through the work that examines the user demand characteristics of the Internet video service, and analyze the challenges faced by fog-based video distribution platforms. We review the work that addresses the challenges of fog-based platforms. We divide such work into two categories: video replication and access problem. Video replication for fog must be distributive, responsive and easy to implement. Uncoordinated replication schemes are simple and distributive, but lack global popularity information. Coordinated replication schemes have the global information but may ignore the local preference of the video popularity. Both wired and wireless networks face the video access problem, namely, to decide where to get the video content for each user. For wired networks, effective solutions manage to coordination of millions of devices. The fog operators usually cluster the user demand or the fog devices into groups to reduce the problem complexity. For wireless networks, current work jointly optimizes the user replication and access problem together by jointly considering the user distribution and the video popularity. We discuss two real implemented and commercially operational fog distributed systems and examine their methods to solve video replication and access problems.

We further go through some future research directions. Combining fog-based video distribution platform with other research topics can provide more interesting and useful applications. With the increasing ubiquitous deployment of fog device, fog computing offers great possibilities for new service and applications. Despite many new directions for fog computing, we briefly overview some future research directions that closely related to video content distributions as follows. Some other emerging directions of fog-based content distribution may include:

- Augmented reality: Augmented reality offers the live view of the combination of real-world environment and computer-generated perceptual (video) information. As the service is location-based, offload the video content to the fog devices near the user location can dramatically reduce the network traffic and the cost. The augmented reality service is closely related to the localization technologies. Both localization and video processing service can share the same fog network infrastructure [11, 49].

- Video data analytics: Massive video data from CCTV monitoring cameras and user generated content from smartphones still keeps increasing [10]. To dig out useful information, we need to analyze such data. It requires a huge amount of storage capacity and network bandwidth to store and transmit such video contents. Fog devices near the video source (e.g., routers and computation power in the camera) can analyze the video directly, or at least preprocess the video to extract features and reduce the data size [2, 45].

- Blockchain-based fog CDN: *Blockchain* offers a fully distributive solution to manage the transactions over the Internet. With blockchain, multiple parities (i.e., content providers, ISP, fog device owner, end users, etc.) can negotiate through this blockchain platform distributively to reach a consensus on how the network shall deliver the video contents (network control) and how the fog device providers be paid for their contribution (monetary transaction) [18,35]. Generally, any databases used in the video service can utilize the blockchain technology, and we can store the databases distributively in the fog devices.

REFERENCES

[1] Baidu. Baidu P2P CDN. http://baijin.baidu.com/.

[2] Flavio Bonomi, Rodolfo Milito, Preethi Natarajan, and Jiang Zhu. *Fog Computing: A Platform for Internet of Things and Analytics*, pages 169–186. Springer International Publishing, Cham, 2014.

[3] Eyuphan Bulut and Boleslaw K. Szymanski. WiFi access point deployment for efficient mobile data offloading. In *Proceedings of the First ACM International Workshop on Practical Issues and Applications in Next Generation Wireless Networks*, PINGEN '12, pages 45–50, New York, NY, USA, 2012. ACM.

[4] Crazy CDN. Crazy CDN. http://www.crazycdn.com/product.html.

[5] M. Cha, H. Kwak, P. Rodriguez, Y. Y. Ahn, and S. Moon. Analyzing the Video Popularity Characteristics of Large-Scale User Generated Content Systems. *IEEEACM Trans. Netw.*, 17(5):1357–1370, October 2009.

[6] Zhangyu Chang and S.-H. Gary Chan. Bucket-Filling: An Asymptotically Optimal VoD Network with Source Coding. *IEEE Trans. Multimed.*, 17(5):723–735, May 2015.

[7] Zhangyu Chang and S.-H. Gary Chan. Video Management and Resource Allocation for a Large-Scale VoD Cloud. *ACM Trans Multimed. Comput Commun Appl*, 12(5s):72:1–72:21, September 2016.

[8] Fei Chen, Haitao Li, Jiangchuan Liu, Bo Li, Ke Xu, and Yuemin Hu. Migrating big video data to cloud: A peer-assisted approach for VoD. *Peer-to-Peer Netw. Appl.*, pages 1–15, July 2017.

[9] Liang Chen, Yipeng Zhou, Mi Jing, and Richard T. B. Ma. Thunder crystal: A novel crowdsourcing-based content distribution platform. In *Proceedings of the 25th ACM*

Workshop on Network and Operating Systems Support for Digital Audio and Video, NOSSDAV '15, pages 43–48, New York, NY, USA, 2015. ACM.

[10] VNI Cisco. Cisco visual networking index: Forecast and methodology 2016–2021.(2017), 2017.

[11] A.V. Dastjerdi, H. Gupta, R.N. Calheiros, S.K. Ghosh, and R. Buyya. Chapter 4 - fog computing: principles, architectures, and?applications. In Rajkumar Buyya and Amir Vahid Dastjerdi, editors, *Internet of Things*, pages 61 – 75. Morgan Kaufmann, 2016.

[12] Prithula Dhungel, Keith W Ross, Moritz Steiner, X Hei, and Y Tian. Measurement study of xunlei: Extended version. Technical report, Technical report, Polytechinc Institute of NYU, 2011.

[13] Prithula Dhungel, Keith W. Ross, Moritz Steiner, Ye Tian, and Xiaojun Hei. Xunlei: Peer-assisted download acceleration on a massive scale. In Nina Taft and Fabio Ricciato, editors, *Passive and Active Measurement*, pages 231–241, Berlin, Heidelberg, 2012. Springer Berlin Heidelberg.

[14] M. Garetto, E. Leonardi, and S. Traverso. Efficient analysis of caching strategies under dynamic content popularity. In *2015 IEEE Conference on Computer Communications (INFOCOM)*, pages 2263–2271, April 2015.

[15] Michele Garetto, Emilio Leonardi, and Valentina Martina. A Unified Approach to the Performance Analysis of Caching Systems. *ACM Trans Model Perform Eval Comput Syst*, 1(3):12:1–12:28, May 2016.

[16] N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire. Wireless video content delivery through distributed caching and peer-to-peer gossiping. In *2011 Conference Record of the Forty Fifth Asilomar Conference on Signals, Systems and Computers (ASILOMAR)*, pages 1177–1180, November 2011.

[17] N. Golrezaei, P. Mansourifard, A. F. Molisch, and A. G. Dimakis. Base-station assisted device-to-device communications for high-throughput wireless video networks. *IEEE Trans. Wirel. Commun.*, 13(7):3665–3676, July 2014.

[18] N. Herbaut and N. Negru. A Model for Collaborative Blockchain-Based Video Delivery Relying on Advanced Network Services Chains. *IEEE Commun. Mag.*, 55(9):70–76, 2017.

[19] H. Hu, Y. Wen, and D. Niyato. Public cloud storage-assisted mobile social video sharing: A supermodular game approach. *IEEE J. Sel. Areas Commun.*, 35(3):545–556, March 2017.

[20] Wen Hu, Zhi Wang, Ming Ma, and Li-Feng Sun. Edge video CDN: A Wi-Fi content hotspot solution. *J. Comput. Sci. Technol.*, 31(6):1072–1086, November 2016.

[21] Cheng Huang, Jin Li, and Keith W. Ross. Can internet video-on-demand be profitable? In *Proceedings of the 2007 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*, SIGCOMM '07, pages 133–144, New York, NY, USA, 2007. ACM.

[22] Fatemeh Jalali, Rob Ayre, Arun Vishwanath, Kerry Hinton, Tansu Alpcan, and Rod Tucker. Energy Consumption of Content Distribution from Nano Data Centers Versus Centralized Data Centers. *SIGMETRICS Perform Eval Rev*, 42(3):49–54, December 2014.

[23] Wenjie Jiang, Stratis Ioannidis, Laurent Massoulié, and Fabio Picconi. Orchestrating massively distributed CDNs. In *Proceedings of the 8th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '12, pages 133–144, New York, NY, USA, 2012. ACM.

[24] Ikhsan Putra Kurniawan, Hidayat Febiansyah, and Jin Baek Kwon. Cost-effective content delivery networks using clouds and nano data centers. In *Ubiquitous Information Technologies and Applications*, Lecture Notes in Electrical Engineering, pages 417–424. Springer, Berlin, Heidelberg, 2014.

[25] Nikolaos Laoutaris, Pablo Rodriguez, and Laurent Massoulie. ECHOS: Edge Capacity Hosting Overlays of Nano Data Centers. *SIGCOMM Comput Commun Rev*, 38(1):51–54, January 2008.

[26] M. Leconte, G. Paschos, L. Gkatzikis, M. Draief, S. Vassilaras, and S. Chouvardas. Placing dynamic content in caches with small population. In *IEEE INFOCOM 2016 - The 35th Annual IEEE International Conference on Computer Communications*, pages 1–9, April 2016.

[27] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo. Pricing and resource allocation via game theory for a small-cell video caching system. *IEEE J. Sel. Areas Commun.*, 34(8):2115–2129, August 2016.

[28] S. Li, I. Doh, and K. Chae. NRIT: Non-redundant indirect trust search algorithm for a cross-domain based CDNi-P2P architecture. In *2018 20th International Conference on Advanced Communication Technology (ICACT)*, pages 992–998, February 2018.

[29] Zhenyu Li, Rong Gu, and Gaogang Xie. Measuring and enhancing the social connectivity of ugc video systems: a case study of youku. In *Proceedings of the Nineteenth International Workshop on Quality of Service*, page 36. IEEE Press, 2011.

[30] Pear Limited. Pear fog CDN. `https://pear.hk/`.

[31] G. Ma, Z. Wang, M. Zhang, J. Ye, M. Chen, and W. Zhu. Understanding performance of edge content caching for mobile video streaming. *IEEE J. Sel. Areas Commun.*, 35(5):1076–1089, May 2017.

[32] M. Ma, Z. Wang, K. Yi, J. Liu, and L. Sun. Joint request balancing and content aggregation in crowdsourced CDN. In *2017 IEEE 37th International Conference on Distributed Computing Systems (ICDCS)*, pages 1178–1188, June 2017.

[33] Ming Ma, Zhi Wang, Ke Su, and Lifeng Sun. Understanding content placement strategies in smartrouter-based peer video CDN. In *Proceedings of the 26th International*

Workshop on Network and Operating Systems Support for Digital Audio and Video, NOSSDAV '16, pages 7:1–7:6, New York, NY, USA, 2016. ACM.

[34] Paolo Medagliani, Stefano Paris, Jérémie Leguay, Lorenzo Maggi, Xue Chuangsong, and Haojun Zhou. Overlay routing for fast video transfers in CDN. *ArXiv Prepr. ArXiv170109011*, 2017.

[35] M. Migliardi, A. Merlo, and A. Passaglia. On the Feasibility of Moderating a Peer-to-Peer CDN System: A Proof-of-Concept Implementation. In *2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC)*, pages 689–694, November 2015.

[36] K.P. Naveen, Laurent Massoulie, Emmanuel Baccelli, Aline Carneiro Viana, and Don Towsley. On the interaction between content caching and request assignment in cellular cache networks. In *Proceedings of the 5th Workshop on All Things Cellular: Operations, Applications and Challenges*, AllThingsCellular '15, pages 37–42, New York, NY, USA, 2015. ACM.

[37] Nitish K. Panigrahy, Jian Li, Faheem Zafari, Don Towsley, and Paul Yu. What, When and Where to Cache: A Unified Optimization Approach. *ArXiv171103941 Cs*, November 2017.

[38] K. Poularakis, G. Iosifidis, A. Argyriou, and L. Tassiulas. Video delivery over heterogeneous cellular networks: Optimizing cost and performance. In *IEEE INFOCOM 2014 - IEEE Conference on Computer Communications*, pages 1078–1086, April 2014.

[39] K. Poularakis, G. Iosifidis, and L. Tassiulas. Approximation algorithms for mobile data caching in small cell networks. *IEEE Trans. Commun.*, 62(10):3665–3677, October 2014.

[40] Thangam Vedagiri Seenivasan and Mark Claypool. CStream: Neighborhood bandwidth aggregation for better video streaming. *Multimed Tools Appl*, 70(1):379–408, May 2014.

[41] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire. Femto-Caching: Wireless content delivery through distributed caching helpers. *IEEE Trans. Inf. Theory*, 59(12):8402–8413, December 2013.

[42] Abdulaziz Shehab, Mohamed Elhoseny, Mohamed Abd El Aziz, and Aboul Ella Hassanien. Efficient Schemes for Playout Latency Reduction in P2P-VoD Systems. In *Advances in Soft Computing and Machine Learning in Image Processing*, Studies in Computational Intelligence, pages 477–495. Springer, Cham, 2018.

[43] Shan-Hsiang Shen and Aditya Akella. An information-aware QoE-centric mobile video cache. In *Proceedings of the 19th Annual International Conference on Mobile Computing & Networking*, MobiCom '13, pages 401–412, New York, NY, USA, 2013. ACM.

[44] L. Sun, M. Ma, W. Hu, H. Pang, and Z. Wang. Beyond 1 million nodes: Crowdsourced video CDN: Architecture, technology, and economy. *IEEE Multimed.*, PP(99):1–1, 2017.

[45] Bo Tang, Zhen Chen, Gerald Hefferman, Tao Wei, Haibo He, and Qing Yang. A hierarchical distributed fog computing architecture for big data analysis in smart cities. In *Proceedings of the ASE BigData SocialInformatics 2015*, ASE BDSI '15, pages 28:1–28:6, New York, NY, USA, 2015. ACM.

[46] Vytautas Valancius, Nikolaos Laoutaris, Laurent Massoulié, Christophe Diot, and Pablo Rodriguez. Greening the Internet with Nano Data Centers. In *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*, CoNEXT '09, pages 37–48, New York, NY, USA, 2009. ACM.

[47] Zhi Wang, Lifeng Sun, Xiangwen Chen, Wenwu Zhu, Jiangchuan Liu, Minghua Chen, and Shiqiang Yang. Propagation-based social-aware replication for social video contents. In *Proceedings of the 20th ACM International Conference on Multimedia*, MM '12, pages 29–38, New York, NY, USA, 2012. ACM.

[48] W. Wu and J. C. S. Lui. Exploring the optimal replication strategy in P2P-VoD systems: Characterization and evaluation. *IEEE Trans. Parallel Distrib. Syst.*, 23(8):1492–1503, August 2012.

[49] S. Yi, Z. Hao, Z. Qin, and Q. Li. Fog computing: Platform and applications. In *2015 Third IEEE Workshop on Hot Topics in Web Systems and Technologies (HotWeb)*, pages 73–78, Nov 2015.

[50] M. Zhanikeev. Fog cloud caching at network edge via local hardware awareness spaces. In *2016 IEEE 36th International Conference on Distributed Computing Systems Workshops (ICDCSW)*, pages 184–188, June 2016.

[51] Y. Zheng, J. Chen, and Y. Cheng. A network adaptive fragmentation algorithm for P2P-based CDN. In *2017 IEEE 9th International Conference on Communication Software and Networks (ICCSN)*, pages 1031–1036, May 2017.

[52] Y. Zhou, T. Z. J. Fu, and D. M. Chiu. Statistical modeling and analysis of P2P replication to support VoD service. In *2011 Proceedings IEEE INFOCOM*, pages 945–953, April 2011.

[53] Y. Zhou, T. Z. J. Fu, and D. M. Chiu. A unifying model and analysis of P2P VoD replication and scheduling. *IEEEACM Trans. Netw.*, 23(4):1163–1175, August 2015.