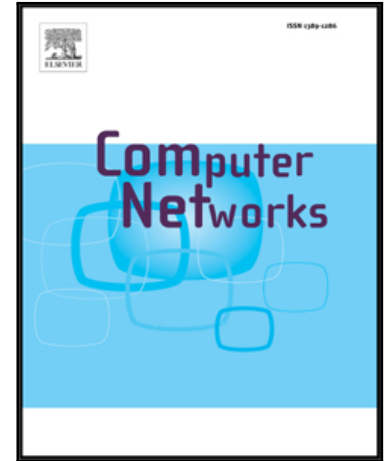# Accepted Manuscript

Survey on Peer-assisted Content Delivery Networks

Nasreen Anjum, Dmytro Karamshuk, Mohammad Shikh-Bahaei,
Nishanth Sastry

Survey on Peer-assisted Content Delivery Networks

Nasreen Anjum[1], Dmytro Karamshuk[2,*], Mohammad Shikh-Bahaei[3,*], Nishanth Sastry[4,*]

Department of Informatics, King's College London

Abstract

Peer-assisted content delivery networks have recently emerged as an economically viable alternative to traditional content delivery approaches: the feasibility studies conducted for several large content providers suggested a remarkable potential of peer-assisted content delivery networks to reduce the burden of user requests on content delivery servers and several commercial peer-assisted deployments have been recently introduced. Yet there are many technical and commercial challenges which question the future of peer-assisted solutions in industrial settings. This includes among others unreliability of peer-to-peer networks, the lack of incentives for peers' participation, and copyright issues. In this paper, we carefully review and systematize this ongoing debate around the future of peer-assisted networks and propose a novel taxonomy to characterize the research and industrial efforts in the area.

To this end, we conduct a comprehensive survey of the last decade in the peer-assisted content delivery research and devise a novel taxonomy to characterize the identified challenges and the respective proposed solutions in the literature. Our survey includes a thorough review of the three very large scale feasibility studies conducted for BBC iPlayer, MSN Video and Conviva, five large commercial peer-assisted CDNs - Kankan, LiveSky, Akamai NetSession, Spotify, Tudou - and a vast scope of technical papers. We focus both on technical challenges in deploying peer-assisted solutions and also on non-technical challenges caused due to heterogeneity in user access patterns and distribution of resources among users as well as commercial feasibility related challenges attributed to the necessity of accounting for the interests and incentives of Internet Service Providers, End-Users and Content Providers. The results of our study suggest that many of technical challenges for implementing peer-assisted content delivery networks on an industrial scale have been already addressed in the literature, whereas a problem of finding economically viable solutions to incentivize participation in peer-assisted schemes remains an open issue to a large extent. Furthermore, the emerging Internet of Things (IoT) is expected to enable expansion of conventional CDNs to a broader network of connected devices through machine to machine communication.

Keywords: Survey, Content Delivery Network, Peer-to Peer Network, Peer-assisted CDN

1. Introduction

Recent years have witnessed tremendous growth in video traffic on the Internet as a result of higher broadband data rates, proliferation in smart handheld devices [24] [95] and affordable unlimited data plans offered by Internet Service Providers [51]. An estimated one-third of all online activities on the Internet is spent watching video according to the recent report [100]. Netflix alone is reportedly streaming over 1 billion hours of video each month which is equivalent to almost 7,200,000 Terabytes of video traffic [37], and this figure is rising constantly. The skyrocketing demand for serving video traffic have questioned the effectiveness of the traditional solution of employing special purpose *Content Delivery Networks* (CDNs), to serve such content. Invented at the turn of the century [96], CDNs now constitute the backbone for serving content [25] [80]. Yet, as several recent studies suggest [60] [109], even CDNs

---

*Corresponding author

Email addresses: nasreen.anjum@kcl.ac.uk (Nasreen Anjum), Dmytro.Karamshuk@kcl.ac.uk (Dmytro Karamshuk), m.sbahaei@kcl.ac.uk (Mohammad Shikh-Bahaei), Nishanth.Sastry@kcl.ac.uk (Nishanth Sastry)

[1]PhD Student at Department of Informatics, King's College London.

[2]Postdoctoral researcher at Department of Informatics, King's College London.

[3]Reader in Communications and Signal Processing at Department of Informatics, King's College London S1.11b Strand, London, WC2R2LS, UK

[4]Senior Lecturer at Department of Informatics King's College London S1.04, Strand, London, WC2R2LS, UK

are being stressed by the demands placed by video users during peak hours.

Thus, there is a prodigious interest in searching for alternative content delivery methods that mitigate the stress on CDNs without losing its core objectives. As a first available solution, CDN operators could (and have) deployed more servers across the globe, in order to maintain a balance among user requests and system services. But this requires major investments both in infrastructure and administrative domain [67]. Recently, an alternative approach has been suggested – to employ peer-to-peer technology (P2P) to assist CDN servers and thereby solve scalability and cost issues of traditional CDNs.

The central idea behind such *peer-assisted content delivery networks* (PA-CDNs)[5] is to combine the benefits of two different technologies for content distribution: traditional server-based CDNs, and P2P networks. Traditional CDNs rely on professionally managed and geographically distributed infrastructure. CDN servers can therefore be expected to be highly reliable and available, and are engineered to provide a high quality of service, often governed by service-level agreements (SLAs) between the CDN provider and the content owners whose content is being distributed by the CDN. However, from an economic perspective, traditional CDNs require significant investments for scaling up, as it requires deployment and management of geographically distributed data centres [54].

Interestingly, scaling up is precisely the strength of P2P content delivery: Early on, it was identified that P2P swarms possess the so-called *self-scaling* property [73] [83] – available capacity increases with the number of users in the swarm, as each user downloading content also adds new capacity by acting as a *server* for other users. However, obtaining content through self-organised P2P swarms has proved to be unreliable due to availability issues [52], because in selfish swarming protocols such as BitTorrent, users leave the swarm after obtaining the item they need, making it difficult to put together complete copies of content. Although replacing selfish self-organising swarming with centrally managed P2P content exchange can improve completion rates considerably [81], for CDN-grade reliability, peers would still need to stick around to allow other users in the swarm to complete, or to maintain a distributed copy of the entire content item across the set of active users in the swarm. PA-CDNs solve this remain-

ing issue very simply: CDN servers essentially operate as a back-up node in the P2P content distribution, and can provide a full copy of the content item (similar to *seeds* in BitTorrent terminology), providing high reliability and guaranteed quality of service.

In short, PA-CDNs work as follows: Whenever possible, i.e., whenever there is sufficient capacity to deliver content in the swarm, peers distribute chunks of content amongst each other (typically using centrally managed swarming techniques [81] [93] [124]). However, when there is insufficient capacity in the swarm (e.g., there are no peers near a user with free upload capacity to deliver the content whilst maintaining QoS guarantees), users are served directly from the CDN servers. Unlike traditional CDNs which need to cater to peak demand, servers in PA-CDNs need only be provisioned for periods when P2P swarming will not be self-sufficient because of very few users.

Thus, PA-CDNs can deliver significant benefits *if adopted widely*. Indeed, *significant traffic savings* (i.e., from 50% to 88% of all consumer traffic) from peer-assistance have been recently reported in feasibility studies for a number of large VOD vendors, e.g., BBC iPlayer [50], Conviva [11], and MSN Video [42]. Moreover, in the recent years PA-CDNs have been deployed by a number of leading commercial CDNs including Akamai [124], ChinaCache [117], and Xunlei [121].

Yet there is a long list of obstacle factors which slow penetration of peer-assisted content delivery in industrial settings. This ranges from the concerns related to the innate unreliability of peer-to-peer networks and challenges in managing distributed P2P swarms to copyright issues and lack of mechanisms to incentivise users' participation. To help the research community to assess the pros and cons of peer assisted solution, this paper aims at conducting a critical extensive survey of the related literature.

Our major contributions are as follows:

- Firstly, we provide an extensive review of a decade-old peer-assisted content delivery literature and then capture the most recent research articles. This includes *measurement studies* which assess the feasibility of peer-assisted solution for three commercial service providers: *BBC iPlayer*, *Conviva* and *MSN Video*, and a number of *modelling papers* focused on solving various technical challenges in PA-CDNs.

- Secondly, we devise a *novel taxonomy* to characterize the challenges for deploying peer-assisted networks in commercial settings. Unlike the previous study [66], we analyze a significantly wider

---

[5]In some literature this approach has also been termed as CDN-P2P Hybrid [67] or hybrid CDN-P2P [39]. In this paper, we will almost exclusively use the term PA-CDNs.

range of obstacle factors, including not only the traditionally discussed technical challenges such as reliability and QoS, but also various other factors, including heterogeneity and scale, and inhomogeneous distribution of resources among users. Furthermore, we scrutinize non-technical factors including *economic and policy* challenges that stem from the lack of incentivizing mechanisms for enabling participation from various stakeholders, e.g. Internet service providers and end-users, as well as the copyright and authentication issues.

- Finally, we survey five large commercial PA-CDNs - *Kankan*, *Livesky*, *Akamia NetSession*, *Spotify* and *Tudou*, and analyze how each of them has tackled the identified challenging factors.

Our findings suggest that although many technical challenges (including mechanisms to eliminate playback latency and deal with peers' inaccessibility behind firewalls) are already solved in the existing commercial peer-assisted deployments, some of the non-technical issues are yet to be addressed. This includes, among others, a need for economically sound mechanisms to incentivize user participation, which is reportedly low in some of the existing systems [124] and is recognized as a significant obstacle factor in the others [50]. Last but not the least, most of the commercial PA-CDN studies have recognized that the effective strategies for managing end-user caches can significantly improve performance of peer-assisted content delivery networks, suggesting that the design of predictive content caching algorithms is a very promising direction for future research.

To the best of our knowledge, this is the first survey focusing on challenges for practical large-scale commercial-grade deployment of PA-CDNs. Our work, with its focus on deployable *peer-assisted* content delivery, is complementary to several survey articles which have focused on traditional P2P-based content delivery [7], or on specific P2P issues such as P2P overlay construction [68], or chunk scheduling [62]. The closest work is that of Lu *et al.* [66], which surveys the design space of PA-CDN architectures, and highlights a fundamental choice between a tightly coupled or loosely coupled model of co-operation between the server-assisted and peer-assisted components of the PA-CDN. Our work departs from Lu *et al.* in its focus on technical *challenges* and deployability issues that would affect all kinds of PA-CDNs. Our work also includes non-technical challenges as a first-class concern, and highlights the need to solve these issues before the promise of PA-CDNs can be realised in practice. We

also provide our vision about the future of research in peer assistance.

The rest of the paper is organized as follows. In Section 2, we provide brief knowledge and background in CDN, P2P and PA-CDN, and explain the challenges of PA-CDN. In Section 3, we introduce our taxonomy for surveying PA-CDN literature. In Section 4, Section 5 and Section 6, we focus on heterogeneity, technical and commercial feasibility challenges, respectively. In Section 7, we survey the existing Commercial PA-CDNs. Finally, in Section 8, we point out main challenges in the PA-CDN that require further investigation.

2. Background

In this section, we first provide brief background in P2P and CDN technologies. No wonder, extensive literature is already available in both domains. However, for the reader's convenience, we will provide background knowledge to cover the fundamentals of both approaches. Further, we discuss the overall architecture of PA-CDN along with the main issues for deploying peer-assisted networks in commercial settings.

2.1. Content Delivery Networks

The user requests in CDNs are served by a *request-routing* system. The request-routing system in CDNs typically includes two basic modules for routing user requests to the most suitable edge server; i) request-routing algorithm and ii) request-routing mechanism [80]. The request-routing algorithms are responsible to select the most appropriate edge server on user requests, while the request-routing mechanism is responsible to direct the user requests to a selected edge server within the CDN.

This entire process is designed to be transparent for the end-users and is most commonly implemented on the DNS level: A special purpose DNS server is programmed to redirect users' requests to the IP address of an appropriate edge server by considering some important parameters such as load on edge server and its distance from the client, network proximity, and user perceived latency [19]. Many popular commercial CDNs such as Akamai [43] and PA-CDNs such as NetSession, Spotify, LiveSky, Tudou and kankan, rely on DNS redirection mechanism.

Content Delivery Network is a complex content distribution system and several issues and decisions are involved in managing and administrating the entire CDN infrastructure, including, where to place edge servers [18], what content to replicate [32] [48] [97],

and on which cluster of servers to copy each piece of content [8]. The literature on resolving these issues is extremely rich. For more information, the interested readers are referred to [79].

Content Delivery Networks have been developed in-house by some Internet giants such as Google, Facebook and Microsoft, but, designated content delivery services are available from companies like Limelight, Akamai, and Level 3. To provide a cheap pay-as-you-go service to a broad variety of customers some CDNs have adopted cloud technologies which became known in the literature as *cloud CDNs* [4]. Last but not least, multiple telecom operators (AT&T, BT, Orange, Telefonica, KPN and Verizon among others) to gain a better control over the data services served to their users have deployed their private *telco-CDNs* [16] [30].

Overall, Cisco has estimated that content delivery network traffic will carry nearly two-thirds of all Internet video traffic by 2020 [24]. Yet, several recent studies have reported that CDNs are being stressed by the demands placed during peak hours [60] [109]. In a search for alternatives, the networking community have considered augmenting traditional CDN architecture with P2P networks [11] [42] [50] [94].

2.2. Peer-to-peer Systems

Peer-to-peer systems offer an alternative to CDNs content distribution architecture. In a CDN architecture, clients' requests are served by dedicated servers. By contrast, in a peer-to-peer system, peers act as both clients and servers and contribute their resources - e.g., storage and network bandwidth - to increase the availability of content items within the groups of dedicated peers, i.e., *content swarms*.

Because users dedicate their resources to serve requests from their peers, a peer-to-peer system can in principle scale arbitrarily without a need of a dedicated infrastructure. However, this makes peer-to-peer systems prone to unpredictable changes in content availability - a phenomenon called *peer churn* - when peers frequently and suddenly leave or join the system due to network failures or their own intent [103].

With a fair participation among peers - i.e., when users agree to share their resources in return of access to the system - the content swarms are said to be *self-scalable* [28] - as an increase in demand for a content item yields an equal increase in the number of the content suppliers [73] [83]. Hence, in contrast to CDNs, peer-to-peer systems allow for considerable savings on infrastructure.

Some peer-to-peer applications may rely on dedicated nodes to control, coordinate and manage content swarm. This structure is referred to as *partially centralized P2P system* [88]. For instance, in the early versions of Skype - one of the first peer-to-peer Internet telephony networks - exploited well connected *Supernodes* to relay the traffic from the peers behind firewalls or Network Address Translators (NATs) [12] [114]. Similarly, PPLive - the largest P2P live streaming service - rely on dedicated *Tracker* nodes to store the information about streaming channels, available video chunks and peers [40]. Some versions of Bit-Torrent - a popular file sharing P2P system - also rely on *Tracker* nodes to distribute bulk software updates or multimedia files [88].

To this date, a multitude of survey articles on P2P technology has been presented. For instance, the authors of [7] present a critical analysis on different design features and infrastructural properties of P2P systems and their influence on non-functional aspects such as scalability, resource management, security, fairness and self-organization. A comprehensive survey of various techniques proposed for structured and un-structured P2P networks has been presented in [68]. Similarly, the authors of [62] provided an overview of different approaches to address chunk scheduling techniques and peering mechanisms.

2.3. Peer-Assisted Content Delivery Networks

PA-CDN is a hybrid architecture which aims to combine the advantages of both - CDNs and P2P systems. On the one hand, PA-CDNs rely on contributions of content delivery resources from the end-users and, therefore, allow to minimize infrastructural costs in comparison to traditional CDNs. On the other hand, PA-CDNs reduce the drawbacks of pure P2P systems by replicating the content in a minimal amount of edge servers, therefore, enabling reliable content distribution with a guaranteed quality of service. In this context, the most striking challenge for the research community is *how to integrate traditional CDNs and P2P and at the same time to benefit from the advantages of both*?

The variety of the hybrid peer-assisted content delivery architectures can be generally divided into two main groups based on the degree of centrality in managing P2P overlay systems.

Centralized architecture: In centralized PA-CDNs new users contact the nearest edge nodes via a request-routing system of the CDN. The edge servers register the peers and store their meta information - such as IP addresses, port of the requested content, etc. - in a centralized database. An edge server then forwards several initial video chunks to the requesting peer followed by the list of randomly selected peers from the content swarm. If the peer is unable to establish a connection

4

with the required number of peers or the connection drops, the edge server sends a refreshed list of peers, or if a video chunk is not received within the playback deadline, the edge server serves the request directly.

Pure centralized PA-CDN architecture has several advantages over pure P2P systems. The quality of services is guaranteed in centralized PA-CDNs through the replicas of initial and missing video segments stored at CDN nodes. Additionally, in a centralized PA-CDN architecture the CDN has control over P2P overlay, therefore, considerably simplifying routines for managing content swarms. A pure centralized architecture has been implemented in the vast majority of commercial PA-CDNs [35] [64] [121] [124].

Decentralized architecture: In decentralized peer-assisted CDNs the peer-to-peer overlay is managed by the designated *Tracker-peers* which are selected, for instance, based on their proximity to the edge server. A peer joining a decentralized peer-assisted CDN contacts a Tracker-peer to obtain a list of active peers. If the Tracker-Peer is unable to provide a desired list of peers, the request will be redirected to another close-by Tracker-peer until a desired video chunk is found. If sufficient number of peers to download the content is not available, the Tracker-peer will request the missing video chunks from an edge node. Decentralized architectures minimize the required CDN infrastructure by delegating most of the swarm managing functions to the users. At the same time, decentralized PA-CDNs are more vulnerable to malicious attacks and are more difficult to manage with respect to their centralized alternatives.

Challenges: The benefits of offloading CDN traffic in peer-assisted content delivery depends entirely on the the number of participated users: Larger and more stable P2P swarms can serve more users, therefore, reducing the load from content delivery nodes. However, the size of a content swarm or scalability in a P2P system is affected by many factors, for example: i) heterogeneity in user access patterns: Users arrive in the system asynchronously and stay for a limited amount of time, ii) heterogeneity in resources: Different users have different Internet connections and different devices both affecting how much and how efficiently content can be shared with others. Thus, it is important to understand how the size and sustainability of a content swarm is affected by peers with weak and unstable resources.

The traditional CDNs possess enough resources to guarantee high quality for high definition videos, while peer-to-peer systems inherently suffer from various problems which impact quality of service, including streaming discontinuity due to node churn, partial participation, and heterogeneity in resources. Not surprisingly, the majority of research efforts in peer-assisted content delivery literature have focused on developing strategies for improving quality of service in terms of reducing startup delay and playback delay [39] [41] [53] [58] [65] [98] [112].

To establish connection between a pair of peers in the network the data has to pass through a wide ranging of middle boxes including firewalls and Network Address Translators (NATs). Indeed, it has been reported in [31] that the presence of middle boxes is a challenging issue. A peer inside the private network can initiate a connection with Peers of public network, but a reverse connection is often complicated by administrative policies [29].

Security and copyrights protection is a critical problem for peer-to-peer networks which are innately vulnerable to all sorts of attacks from malicious peers. For example they may upload and share corrupted, malicious or illegal contents in the system or block nodes from using the service [89]. Indeed, pollution of the system with corrupted and unauthorized contents has been named as a serious problems in peer-to-peer networks [105]. Privacy is yet another issue in decentralized P2P systems: private information of the users including their IP addresses, geographic locations and viewing preferences might be exposed to unauthorized users. It is thus important to understand how hybrid peer-assisted content delivery networks can address copyright and security challenges common for P2P networks.

It is equally important to address economic incentives for participating parties: Effectiveness of a PA-CDN deployment depends upon the users' willingness to share their resources but also on ISPs' attitude towards peer-to-peer traffic. The measurement studies of the commercial PA-CDNs - which we discuss in details in section 4 - report a low participation level among users and, so, finding economically viable mechanisms to incentivise user participation is an active research area.

A controversial attitude of ISPs towards P2P networks is yet another challenge for commercial PA-CDN vendors. Historically peer-to-peer systems have been ISP-oblivious and could generate significant amount of cross-ISP traffic for what the ISPs can be charged. The loss of revenue among ISPs has started a tussle between ISPs and P2P systems resulting in blocking and discouraging P2P traffic by the former [74] [75]. Achieving ISP-friendly peer-to-peer traffic is thus a major challenge for commercial deployment of PA-CDNs.
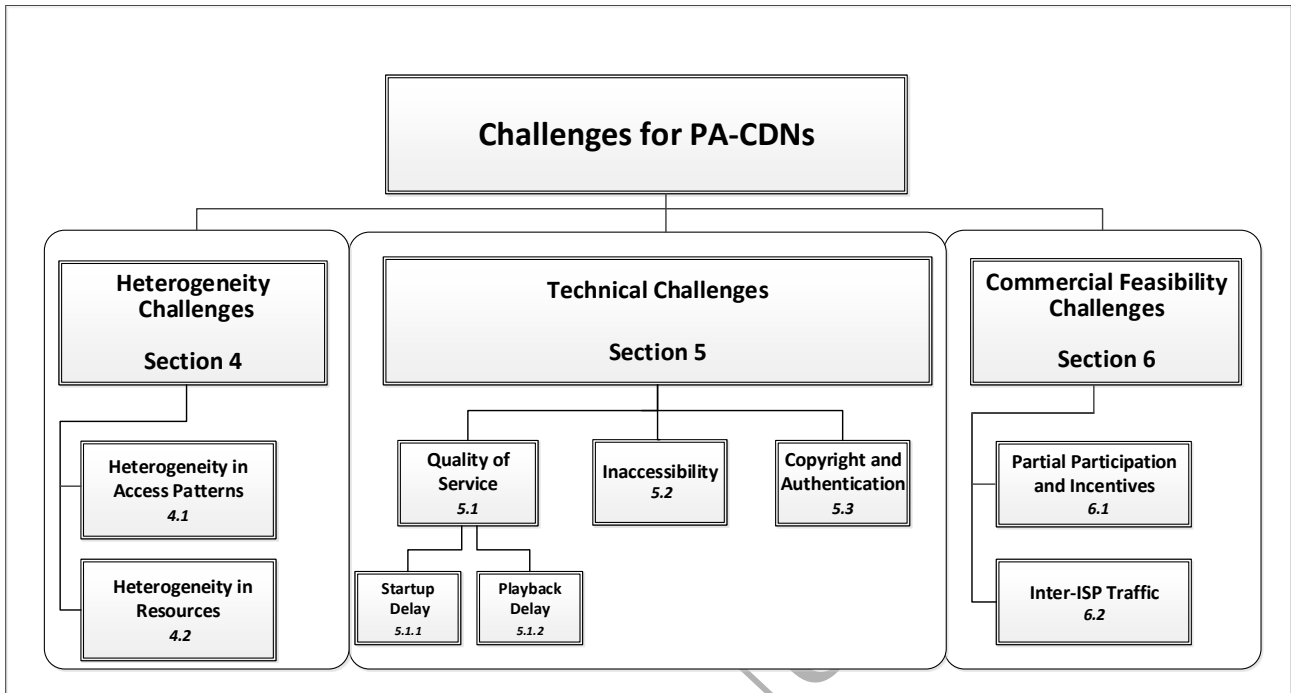
Figure 1: **Taxonomy of Challenges in PA-CDN Literature.**

3. Introduction to Taxonomy for Surveying PA-CDNs

In this section we introduce a new taxonomy for surveying peer-assisted content delivery research. Our survey of PA-CDNs is different from the existing P2P and CDN surveys in several aspects. Firstly, we focus on the challenges specific for hybrid PA-CDNs rather than traditional P2P networks. Although we admit that many of the challenges are common between these two types of networks, in this paper we only consider the approaches which take advantage of the hybrid setup where peers are augmenting the server facilities provided by content delivery networks rather than replacing them. Secondly, we introduce a new taxonomy for surveying the literature on peer-assisted content delivery based on the challenges which peer-to-peer augmentation of traditional CDNs induce for commercial content delivery providers. We characterize each challenge by its origin and distinguish between the challenges caused by particularities in user access patterns, by technical constraints or by ISPs' and end-users' attitude towards peer-to-peer content sharing.

To this end, we identify the main challenges of peer-assistance which have been addressed in the literature and organize them into three different groups. The proposed taxonomy is summarized in Figure 1. The group

of **heterogeneity challenges** characterizes the impact of inhomogeneous user access patterns and resources (e.g., asynchronous arrivals, instant departures, limited upload bandwidth) on the capacity of peer-to-peer swarms and traffic savings achieved from peer-assistance. Many solutions have been proposed to deal with heterogeneity challenges which we will discuss in section 4. In the **technical challenges** we focus on those technical aspects which have been predominantly emphasized in the literature and organize them in the following groups : *Quality of Service* challenges which discuss to which extent the quality of service is affected by the consequences of distributing content delivery responsibilities among often unreliable peers (see subsection 5.1); challenges related to *inaccessibility of peers* behind firewalls and NATs (see subsection 5.2); and challenges related to *copyrights protection* (see subsection 5.3). The group of **commercial feasibility challenges** addresses challenges of the economic incentives for participating parties (see subsection 6.1) and the ISPs' attitude towards peer-to-peer traffic (see subsection 6.2).

4. Heterogeneity Challenges

In this section we will focus on the heterogeneity nuances specific for PA-CDNs. First we will discuss how

6

Table 1: Comparison of Techniques for Heterogeneity Challenges in PA-CDNs

	Year	Service Type	Heterogeneity Challenges	Contributions	Validation Method
[11]	2013	VOD and Live	Heterogeneity in Access Pattern	Completion Ratio: Reported an early quitter phenomenon when users terminate a video session before completion.	Trace Driven Simulation
PowerInfo [119]	2006	VOD	Heterogeneity in Access Pattern	Completion Ratio: Reported a 70% abandon rate among mobile users when viewers quit video sessions after watching first 20 minutes of a video.	Trace Driven Simulation
[51]	2015	VOD	Heterogeneity in Access Pattern	Completion Ratio: A comparison analysis between the abandon ratio among mobile and fixed-line users in a large-scale study of a large catch-up TV platform (BBC iPlayer).	Trace Driven Simulation
Spotify [120]	2013	On-demand Music	Heterogeneity in Access Pattern	Completion Ratio : A low completion ratio among users of Spotify is reported.	Trace Driven Simulation
[50] [51]	2015	VOD	Heterogeneity in Access Pattern	Asynchronicity in user access: Reported temporal concentration of video popularity in a large catch-up TV platform (BBC iPlayer).	Trace Driven Simulation
Spotify [120]	2013	On-demand Music	Heterogeneity in Access Pattern	Asynchronicity in user access: Reported a vivid diurnal fluctuation in users arrivals in Spotify.	Trace Driven Simulation
[42] [50] [108]	2013	VOD	Heterogeneity in Access Pattern	Distribution of content popularity: Reported a skewed distribution of video popularity in MSN video, BBC iPlayer and online social networks (OSNs) respectively.	Trace Driven Simulations
[42]	2007	VOD	Heterogeneity in Access Pattern	Asynchronicity in user access: A large scale study of peer-assisted benefits in MSN Video.	Trace Driven Simulation
Thunder-Crystal [20]	2015	VOD	Heterogeneity in Resources	Peer-churn: Proposed a Thunder-Crystal approach that allows peers to download content items from the cache of smart access points (APs), which are permanently available online.	Simulation
Youku [69]	2016	VOD	Heterogeneity in Resources	Peer-churn: Proposed to address peer-churn problem by deploying dedicated intelligent peer routers on the end-user side.	Trace Driven Simulation
P2PTube [78] [38] [72]	2013 2016 2016	VOD VOD VOD	Heterogeneity in Resources	Peer-churn: Address the problem of peer churn in a peer-assisted design by focusing on peer-to-peer communication between the set-top-boxes which are often permanently powered (even if in sleep mode), permanently connected, and have sufficiently large storage capacities.	Trace Driven Simulation Analytical Model and Simulation Simulation
[42]	2007	VOD	Heterogeneity in Resources	Limited upload bandwidth: Proposed two peer selection policies, Water Leveling (WL), and Greedy Policy (GP) to achieve uniform distribution of the surplus bandwidth among users.	Trace Driven Simulation
[111]	2008	VOD	Heterogeneity in Resources	Locality awareness: Proposed a centralized PA-CDN architecture by employing a DHT-based indexing algorithm *Kadmelia* to select a nearest neighboring peer.	Simulation (PeerSim)

the size of a content swarm or scalability in a P2P system is affected by heterogeneity in user access patterns, then we will discuss how the size and sustainability of a content swarm are affected by peers with weak and unstable resources.

The impact of user heterogeneity on the performance of traditional P2P systems has been analyzed in gory details in the previous literature [13] [40] [103] [116] [119]. While the detailed discussion of these approaches is beyond the scope of this paper, the measurement studies considered in this section have suggested remarkable benefits for traffic offloading in PA-CDNs (i.e., from 50% to 88% of CDN traffic can potentially be offloaded to peers during peak hours) inspite of heterogeneity of user resources and access pat-

terns. Our goal is to understand the factors which enable this remarkable potential of peer-assistance for large content providers, such as BBC iPlayer, MSN Video, and Spotify. Table 1 provides a summary of our analysis for heterogeneity challenges.

4.1. Heterogeneity in Access Patterns

One obstacle in scaling peer-assisted content delivery lies in the way users arrive in and leave the system. With live streaming, for instance, user accesses are innately *synchronized* as the same content is delivered to all users simultaneously. Intuitively, even an unpopular content with synchronized accesses may still form sufficiently large content swarms. In contrast, on-demand accesses might be highly spread in time;

therefore, decreasing the chances of forming sustainable content swarms. The time which users spend in the system (*session duration*) - and hence are available to re-distribute the content with the other peers - varies across different systems, too. For instance, a low *completion ratio*, when users abandon watching video after the first few seconds, has been reported in many previous studies [11] [33] [42] [57] [119]. We review the impact of these factors on the performance of PA-CDNs in the rest of this section.

Completion ratio: In [11] the authors have analyzed the user trace collected from the Conviva media platform and reported a very low completion ratio among users, when they abandon sessions after watching first few chunks. This phenomenon was explained by frequent drop-outs of users on unreliable Internet connections, on the one hand, and a vast variety of choices in video content, on the other hand. More specifically, it was reported that during an average video-on-demand session a user will only complete watching a mere 10% of a content. An average reported completion ratio for the live content was slightly higher, i.e., around 20%, but at the same time the behavior of live users was less predictable as they might join or leave sessions at various times, e.g. to check the score of a football match. Similarly, a measurement study of PowerInfo [119] - a video-on-demand system deployed by the world's largest mobile phone operator (China Mobile) - has reported a 70% abandon rate among users as measured by the fraction of sessions which were abandoned after first 20 mins. A comparison between the abandon rate of mobile and fixed-line users has been conducted in a study of BBC iPlayer accesses. The results suggested that mobile users abandon sessions with a higher rate, i.e., only around 30% of mobile sessions last for longer than a half of a content's duration in comparison to around 50% for the fixed-line sessions [51]. The accesses of Spotify users [120] feature a vivid diurnal pattern: The lengths of user sessions are the longest during morning hours and are gradually decreasing by the end of the day. The user sessions are also shorter during weekends than during working days, and on mobile devices than on desktop computers[120].

Asynchronicity in user accesses: User accesses in live streaming are innately synchronized - i.e., all users watch the same part of the content item simultaneously. Consequently, synchronization in live streaming increases opportunities for peer-to-peer content sharing even for unpopular shows because all accesses are concentrated in time. More surprisingly, many studies have reported a very high temporal concentration of accesses also for the video-on-demand accesses which might in principle be very asynchronous. This phenomenon can be explained by the typical diurnal patterns in user accesses, when most of the users come online in the evening peak hours [11], but also by the content availability policies specific for some video on-demand websites: for example, in catch-up TV systems, such as BBC iPlayer [51], the content items are typically released for a limited amount of time, e.g., 7 to 30 days, and feature a burst of accesses in the first few hours after the release[50]. The diurnal fluctuations in user arrivals are also vivid in Spotify where desktop accesses peak during evening hours, i.e., 6-9pm, although the lengths of evening sessions are considerably shorter than the morning ones[120].

Distribution of content popularity: Along with the temporal concentration of user accesses, many previous studies have reported a skewed distribution of user interest toward small fractions of content items. Indeed, in [50], the authors reported that the Top-10% popular videos account for up to 80% of the total traffic in BBC iPlayer. Similarly, in [42] is reported that the Top-10 videos constituted a significant fraction of CDN traffic in MSN Video. The authors in [108] reported that, video popularity in online social networks (OSNs) follows the Zipf-like distribution. However, the content popularity would chang quite dynamically, for example, due to daily releases of highly popular news and business-related shows [42].

Length of video content: The length of the popular video content matters, too. It has been reported in [42] that, for small size MSN videos, users generally opt to view the entire or most of the video clip, and only 20% of users watch 60% of video content with the length greater than 30 minutes. Moreover, a large fraction of users (i.e., 60%), watch videos without interactions (e.g., stop, forward, rewind etc.), whereas this fraction increases to 80% for videos shorter than 30 minutes. A different pattern is observed for BBC iPlayer - a catch-up TV platform - where most of accesses happen for the 30 or 60 minutes long episodes, typical for TV content.

Feasibility studies: Different feasibility studies [11] [42] [50] have been conducted to investigate the patterns of user accesses and their impact on the performance of PA-CDNs (see Table 2 for details). Balachandran et al. in [11] reported that, despite heterogeneity in the user access patterns and resources, peer-

Table 2

Summary of the Feasibility Studies

Service Provider	Service Type	DataSet			Potential Benefits
		Duration	No. of Sessions	No. of Users	
Conviva.com [11]	VOD and Live	Two Months	30M	8.5M	87% (VOD, Peak hours) 30%(VOD, early quitters) 60% (Live, early quitters)
BBC iPlayer [50]	VOD	One Month	15.9M	2.2M	88% (Despite of all obstacles)
MSN Video [42]	VOD	Nine Months	520M	–	50% (ISP-Friendly PA-CDN) 79Mbps server bandwidth (Greedy Pre-Fetching Policy)

Summary of the feasibility studies conducted for the three large video-on-demand vendors indicating *Potential Benefits* from peer-assisted, *Duration* of study and the scale of the system, i.e., *Number of sessions* and *Number of Users*.

assistance has a potential to reduce the burden of CDN servers in *Conviva* media platform for up to 87% during peak hours. Similarly, in [50] the authors built on observation that the users of BBC iPlayer spend considerable amount of time online while watching long duration multimedia contents (e.g., TV shows and feature movies) and suggested that up to 88% of BBC iPlayer traffic can be offloaded to peers. Simulation results from [11] have shown that up to 30% of traffic could be offloaded to P2P system by taking advantage of bootstrapping, i.e., downloading initial chunks from multiple sources to decrease the startup delay. It has been also shown that the savings from bootstrapping are typically higher in live streaming, where CDN could offload up to 60% traffic due to the fact that peers stay longer online and have more content items to upload. The authors of [42] reported that, by employing bandwidth saving mechanism, the server bandwidth requirements could be reduced from 2.20 Gbps to 79 Mbps (for details see-Limited upload bandwidth- in "page 10").

4.2. Heterogeneity in Resources

Heterogeneity in resources is yet another obstacle in implementing peer-assisted content delivery: Different users have *different Internet connections* and *different resource capacities* - both affecting how much and how efficiently content can be shared with other peers. Thus, it is important to understand how the size and sustainability of a content swarm is affected by peers with weak and unstable resources. For example, some peers may disconnect frequently, due to hardware or network failures; therefore, contributing a little to serving requests from others. Moreover, peers with higher bandwidth will download faster and quickly

leave the system, thus, reducing the opportunities for content re-sharing. The asymmetry between upload and download bandwidths of end-mile Internet connections is yet another obstacle for video streaming in many peer-to-peer applications such as Bit-torrent, PPlive, and Maze networks: A downstream rate of ADSL2+, for example, can be as much as 17 times higher than the upstream rate, suggesting a potential bottleneck in uploading bandwidth when streaming a high resolution content item from peers [10].

Peers churn: There are a number of other factors apart from user interests - discussed in the previous section - which impact the dynamics of peers' availability, or *peers churn*, in peer-to-peer systems: A user session might be interrupted due to a network failure, overloaded CPU or a software crash [55]. A downloading user affected by a sudden departure of an uploading peer must be dynamically re-assigned to other available peers or edge servers, at the same time, minimizing the negative effect of churn on users' watching experience. To tackle the negative impact of peers churn on the performance of peer-assisted content delivery systems, the authors in [20] proposed a crowdsourcing-based content distribution system called Thunder Crystal. The Thunder Crystal approach relies on the smart access points (APs) which are assumed to be widespread among users in future. A smart AP is presumed to be equipped with a large storage for caching high resolution videos and is meant to share the content with the peers in exchange of some benefits obtained from the content provider, e.g., discounted subscription fees. Similarly, Youku, one of the largest VOD service in China, has deployed millions of dedicated intelligent peer routers with 8GB storage

9

capacity at consumers' homes and offices to assist content distribution [69]. Based on a trace driven evaluation of the YouTube traffic, the authors in [78] proposed a P2PTube approach which relies on exploiting set-top-boxes to assist content distribution. As the results from the [78] suggest, as much as 46% of videos can be served from peers in the proposed system - an advance of 10% with respect to the previous result reported in [125]. The authors in [38] devised a mathematical model to evaluate the impact of peer-churn on a PA-CDN, when the users of set-top-boxes are not willing to share their resources. Similarly, to deal with peer churn the authors in [72] proposed Home Box-assisted approach which relies on exploiting set-top-boxes as proxies between users and CDNs.

Limited upload bandwidth: The impact of the limited upload bandwidth on peer-assisted content distribution has been evaluated in [42]. The authors demonstrated that although the demand for a high resolution video content is growing over time, its growth rate is in fact smaller than that of the average upload bandwidth among Internet users. This indicates a very positive trend for the future of peer-assisted content delivery, suggesting that the upload capacities of the end-user Internet connections in future would allow for pervasive peer-to-peer video sharing. To improve the quality of service without putting too much of a burden on the peers, the authors of [42] also suggest two different peer selection policies: Water Levelling (WL) and Greedy Policy (GP). With the WL approach the surplus uploading bandwidth is uniformly distributed among all peers, while in the GP approach each user simply dedicates its remaining upload bandwidth to the nearest user. The simulation results show if a peer-assisted solution was deployed using a greedy policy, the server bandwidth requirements could be reduced from 2.20 Gbps to 79 Mbps.

Locality awareness: The topology of all P2P systems is independent of underlying physical network as they are built upon application level overlays. A biased selection of peers without considering underlying physical topology might lead to severe performance degradation in terms of access delays and bandwidth wastage if, for instance, peers located within the same building use two different ISPs and so, although physically placed close to each other, are very distant in network terms [5]. Early PA-CDN approaches have been network-oblivious, i.e., peer's locality was not taken into consideration, therefore, creating network overhead when delivering from distant nodes, many of the recent papers have addressed this problem in detail. The au-

thors of [111] proposed to limit the P2P traffic within sub-networks, or behind common gateways. The authors exploited the modified version of *Kademlia distributed hash tables* (DHT) to conduct searches for the closest peers. Kademlia creates and stores a 160 bit long hashed ID of each node when he joins the system. The node ID is composed of a 32 bit-long IP address and 128 randomly generated bits. Based on the IP addresses and key values stored in the DHT, Kademlia evaluates the distance between a pair of peers. Similar idea of using Kademlia algorithm has also been suggested in [9], whereas the authors of [27] have proposed a locality aware mechanism based on the website content.

5. Technical Challenges

In this section we focus on the technical aspects of implementing peer-assisted content delivery. First, we consider different approaches which have addressed QoS aspects of peer-assisted content delivery and organize them into two groups of papers which focused on either reducing *startup latency*, occurring at the start of a video streaming session, or *playback latency* occurring throughout the session. Then we discuss how copyright and security, and peers inaccessibility problem have been addressed in the literature. Table 3 provides a summary of our analysis for technical challenges.

5.1. Quality of Service (QoS)

One of the early criticisms to the idea of delegating content delivery functions to peers was in their unreliable behavior in comparison to traditional content delivery networks and - as a consequence - their inability to provide required level of quality of service for on-line video streaming. In contrast to CDNs, peer-to-peer systems inherently suffer from various problems which impact quality of service, including node churn, partial participation, heterogeneity in resources and users' access pattern. The delivery of high quality multimedia contents to the customers is the challenging issue for many network and content delivery services [14] [62] [79] [15].

5.1.1. Startup Delay

To achieve a lower startup delay, the authors in [112] proposed a three phase streaming hybrid CDN-P2P architecture that allows peers to download initial chunks of a content item from the geographically closed CDN nodes and remaining chunks from a P2P swarm. The

Table 3: Comparison of Techniques for Solving Technical Challenges in PA-CDN

	Year	Service Type	Technical Challenges	Contributions	Validation Method
[112]	2006	Live	Startup Delay (QoS)	Proposed to download initial chunks of a content item from the geographically closed CDN nodes.	Analytical Model and Simulation Study
[39]	2008	Live	Startup Delay (QoS)	Proposed an effective buffer management technique that divides a client's buffer into a *CDN region* (served from CDN nodes) and a *peer region* (served from neighboring nodes).	Simulation (PeerSim)
[65]	2011	Live	Startup Delay (QoS)	Proposed to divide the playback buffer into a *CDN region*, a *peer region* and an *emergency region*. A *peer region* is converted into an *emergency region* if a user fails to download required chunks from neighboring peers before the playback deadline.	Simulation
[41]	2012	Live	Startup Delay (QoS)	Proposed a P2P resource scheduling algorithm to achieve lower startup delays and continuous transmission.	Simulation
SocialTube [58] [98]	2014	VOD	Startup Delay (QoS)	Proposed a chunk pre-fetching algorithm based on social relationships and physical locations of peers in Online Social Networks.	Simulation
G-CP2P [53]	2011	—	Playback Delay (QoS)	Proposed a location-based peer selection algorithm by employing a CAN [84] technique that clusters peers based on the latency and the channel they are watching.	Simulation (PeerSim)
[101] [102]	2004	Live	Inaccessibility	Proposed a centralized hybrid CDN architecture based on the empirical study of BitTorrent to address the problems of firewalls.	Analytical Model and Simulation Study
[115]	2008	Live	Copyright and Authentication	Proposed a peer authorization mechanism by employing a network coding scheme to prevent illegal distribution of the content in the system.	Simulation
[49]	2010	VOD	Copyright and Authentication	Proposed a centralized PA-CDN architecture that authorizes content items via a Tracker Server.	Analytical Model

authors of [39] [65] proposed a strategy for improving startup delay via an effective buffer management on a peer's side. Particularly, Ha et al. in [39] suggested that, for minimizing startup latency the buffer's part at the start of the playback must be filled in with a high priority. In the proposed schema, the playback is divided into two regions such that a smaller region close to the start of the playback is served from CDN nodes and the rest of the playback is served from peers. The relative size of the two regions depends on the number of CDN nodes available in the system and on the number of participating peers. A newly arrived peer will firstly approach the edge server to fulfill the *priority region* fast, and then the rest of video chunks will be downloaded from the peers.

Similarly, Lu et al. in [65] suggested organizing the playback buffer into three different regions, where a *startup region* and a *common region* are equivalent to the ones proposed in [39]. However, an additional *emergency region* is introduced in case a user fails to download required chunks from peers before the playback deadline. The schema also proposes a fairness control mechanism when the upload bandwidth of peers is firstly used to serve *superior peers* who are more sta-

ble, spent more time online or who have shared a good portion of their upload bandwidth before. The simulations conducted by the authors of this proposed approach have shown an improvement of 2.5 sec in the average startup latency in comparison to the traditional peer-to-peer networks.

The authors of [41] suggested that a lower startup delay and continuous transmission could be achieved by utilizing P2P resources more efficiently. The core of this proposal lies in employing a resource scheduling schema which is able to handle emergent video chunks before their playback deadlines occur. The schema relies on the three types of queues for video chunk requests. When a peer first enters the system and issues a request for a content item, its request will be put in the first queue based on the first-come-first-serve discipline (FCFS). The edge server will provide few initial video chunks to peers from the first queue for immediate playback, whereas to obtain the rest of the video a user will be redirected to a subset of neighbouring active peers. If a peer cannot receive a desired chunk from peers before the playback deadline, its request for an urgent chunk will be put in the second queue based on the *earliest playback deadline first discipline*. In case a request for

11

a video chunk can only be served from an edge server, it will be scheduled in the third queue with FCFS discipline. The simulation results showed that in comparison to the previous approaches, the proposed scheduling mechanism can reduce the average startup delay from 21.36 seconds to 20.45 seconds (i.e., an improvement of 4.26%).

The authors of [58] [98] exploit the *social ties* between users and the *locality of interests* to assist peer-assisted sharing of content in Facebook. Since, users are more likely to check-out the content shared by their friends; the system pre-fetches few initial chunks of the video shared by a user's friends in her cache before she even requests the content. When a request is actually made, the first initial chunks are played from a local cache, therefore, dramatically reducing the startup delay. The authors suggest to group peers with similar interests and physically close to each other in the interest-based clusters (a peer can become a part of multiple clusters) and download all remaining chunks from the neighboring peers in the cluster. The simulation results suggest that this approach can reduce the startup delay by up to 2 sec in comparison to the previous results in [22] [42].

5.1.2. Playback Delay

Apart from the time it takes to start playback (i.e., startup latency) it is also important to optimize the throughput during a user's session to ensure the playback is not interrupted for buffering. Different solutions to minimize the playback latency, e.g., the time it takes for a data packet to reach a downloading peer from an uploading peer during a user's session, have been proposed in the literature. Thus, the authors of [53] proposed a decentralized PA-CDN architecture - named the Group Based CDN-P2P (G-CP2P) - which estimates the latency between two peers using a set of *landmarks*. In the G-CP2P strategy, peers join different P2P groups based on their latency and a Live TV channel they are watching. Each group is controlled and managed by a special peer called the *Super Peer* (SP) which is selected based on the distance from the edge server. An SP exploits a distributed hash tables algorithm (DHT) called a content addressable network (CAN) [84]. The algorithm is based on the binning technique proposed in [85] and operates as follows. When a peer is selected as an SP, its round trip time (RTT) is used as a landmark or a key such that all SPs are arranged and partitioned in increasing order of RTT in a CAN zone and form separate clusters. Each new peer joins an SP which has the same order of RTT and forms a group with the same cluster zone. Thanks to this locality awareness

mechanism the proposed solution has showed an average decrease of 500 ms in the playback delay with respect to the previous location oblivious result in [112].

In summary, our analysis in this section suggests that, the most promising compromise in mixing advantages of peer-assisted and traditional CDNs to improve QoS has been so far found in bootstrapping video streaming with initial chunks downloaded from CDN nodes and delegating the remaining work to peers [39] [65]. Similarly, the peer-assisted content delivery system may benefit from redirecting streaming requests to CDNs in emergency cases when no sufficient upload bandwidth is available among the peers to meet playback deadlines [112]. As we have discussed before, many of these approaches have been already employed by the existing commercial PA-CDNs.

5.2. Inaccessibility

The peers inaccessibility problem; when a peer inside a private network can initiate a connection with the peers of public networks, but a reverse connection is often complicated by administrative policies [29], has been discussed in some of the earliest partially centralized PA-CDN articles [101] [102]. The authors proposed to provide accessibility within local area networks (LANs). When a peer joins the system and sends a request to a Tracker for the list of uploading peers, the Tracker Server will attempt to make a connection in reverse. If successful, it will add the peer into its list of accessible peers, thus, forming a cluster of nearby peers able to receive incoming connections. The simulation results in [102] suggest that, the users which are connected outside a firewall have slower downloading rates since they are unable to establish connections with the nodes in private networks, whereas peers behind firewalls exchange more, therefore, leading to an increase in downloading speed.

Many other solutions have been proposed to deal with inaccessibility issues in various perspectives and situations including STUN (Simple Traversal of User Datagram Protocol through Network Address Translators) [91], TURN (Traversal Using Relay NAT) [70], RSIP (Realm Specific IP) [17], MIDCOM (Middle box Communication) [99], UPnP (Universal Plug and Play), ICE (Interactive Connectivity Establishment) [90], etc. It is worth noting, that all currently deployed NAT traversal solutions have their limitations (for instance - STUN does not support any feedback mechanism) and their further development remains an interesting research direction [21].

12

5.3. Copyright and Authentication

Preserving copyrights whilst allowing users to store and share content items is yet another challenge for peer-assistance. To address this concern the authors of [115] proposed a peers authorization mechanism and a network coding scheme in which each packet is encoded and decoded at the node level using efficient linear codes, thus, allowing for copyrights protection. Before downloading a content item from the nearest edge server, peers must be authorized by a content publisher, on the one hand, and use a network encoding scheme to protect the content from unauthorized accesses, on the other. The authors of [49] proposed a control schema over copyrights at a Tracker server in which only legally authorized content items are distributed to the peers.

6. Commercial Feasibility Challenges

In this section, we analyze the commercial feasibility challenges in implementing peer-assisted content delivery systems. We focus on the mechanism which incentives different counterparts, i.e., ISPs, content providers and end-users, to participate not only in peer-assisted content delivery but also on the obstacle factors which restrict content providers and ISPs to rely on peer-to-peer communications. Table 4 provides a summary of our analysis for commercial feasibility challenges.

6.1. Partial Participation and Incentives

As we have seen in the previous sections, the PA-CDN approach can drastically reduce the bandwidth cost for CDNs. However, the driving force of peer-assisted content distribution is the users and so the performance of peer-assistance entirely depends on users' willingness to participate. Therefore, it is similarly important to motivate end-users to share their upload bandwidth. Indeed, it has been shown in [124] that only around 30% of all peers opt-in to contribute their resources in Akamai NetSession, whereas in [50] the authors named peers' participation as the main obstacle factor which affects the amount of bandwidth savings achievable with peer-assistance. The influence of the peer co-operation on the scaling properties of peer-assisted CDNs has been analytically studied in [38] [47] and [46]. While the results from [38] [47] have indicated non-cooperative peers as a major problem for scaling peer-assisted CDNs, the findings from the latter work emphasize on the opposite - PA-CDNs with co-operative peers are extremely scalable. Overall, these results suggest that finding economically viable solutions to incentivize peer participation is a crucial

factor for implementing peer-assisted solution in commercial settings.

Incentive policies: Various approaches have been proposed to incentivise user participation and discourage *free riding*. A *limited contribution policy* presented in [112] obliges every user to contribute some fraction of its upload bandwidth resources to a limited number of sessions, for a limited period of time or both. In [34] the authors present an economic model for PA-CDNs in which user participation in peer-assistance is incentivized via free high quality video offers. In return, each subscribed peer is required to sign an SLA (service level agreement) with the content provider on the amount of upload bandwidth he/she is obliged to contribute to the system. The authors in [101] [102] proposed a policy akin to Bit-Torrent system to incentivize peers based on their stability and resources contributed to the system. The authors in [65] proposed an incentive policy that selects peers directly via edge nodes based on the bandwidth they contribute to the system. Similarly, authors in [26] introduced a superiority index for peers that are more stable and share higher bandwidth.

The authors of [76] have proposed a peer-assisted model with economic incentives for all participating parties including both peers and ISPs. It is suggested that ISPs can reduce the inter-ISP transit costs by localizing P2P traffic within its domain. In its turn, ISPs can motivate end-users to sell their idle resources by offering virtual money or discounts in their Internet bills, and naturally, a CDN can offload significant amount of traffic to participating peers and, therefore, reduce its bandwidth cost. Jeong-woo et al. [23] presented a *cooperative game theory* approach to validate a profit sharing mechanism with multiple content providers and peers. The framework proposed in the paper provides an analytical model to analyze the fairness assessment of all players' contributions if they agree to cooperate. The results of the analysis suggest that the peer-assisted system with high level of co-operation among peers and multiple providers can yield aggregate profit to all participating players.

The Qtrade system described in [110] provides a QoE-aware incentive policy which considers the visual quality of videos uploaded by the peers as a criterion for calculating their contributions. The proposed system employs the video quality metric (VQM) [82] for measuring the loss in users' perceived QoE and distributes the surplus bandwidth to the top-scoring peers.

Table 4: Comparison of Techniques for Commercial Feasibility Challenges in PA-CDNs

	Year	Service Type	Commercial Feasibility Challenge	Contributions	Validation Method
[112]	2006	Live	Partial Participation and Incentives	Proposed a limited-incentive policy where every peer is bound to contribute his resources for a limited amount of time or limited number of sessions.	Analytical Model and Simulation Study
[34]	2015	—	Partial Participation and Incentives	Present an economic model for PA-CDNs in which user participation in peer-assistance is incentivized via free high quality video offer.	Analytical Model and Simulation Study
[101] [102] [65] [26]	2004 2011 2013	Live Live VoD	Partial Participation and Incentives	Presented various incentive polices in which most stable peers or peers who contribute the most are directly serviced by the edge servers.	Analytical Model and Simulation Study Simulation Analytical Model and Simulation Study
MPCDN [76]	2009	VOD	Partial Participation and Incentives	Proposed a business model that encourages ISPs to motivate their users to share their upload bandwidth.	Analytical Model and Simulation Study
[23]	2014	VOD	Partial Participation and Incentives	Proposed an analytical model to asses the fairness and incentives for ISPs, end-users and content providers.	Analytical Model
Qtrade [110]	2015	VOD	Partial Participation and Incentives	Proposed a QoE-aware incentive policy that motivates peers to share high bandwidth content.	Simulation (PeerfactSim KOM)
CPH-VOD [67]	2010	VOD	Inter-ISP Traffic	Proposed to deploy special purpose super nodes (SN) within each ISP to coordinate co-operation between CDN and P2P overlays.	Simulation
[78]	2013	VOD	Inter-ISP Traffic	Proposed to employ a DHT-based indexing algorithm Kadmelia and ALTO protocols to minimize the inter-ISP traffic in large ISPs.	Simulation
Nextshare [9]	2012	Live	Inter-ISP Traffic	Proposed to deploy a layer of proxy CDN nodes within each ISP network, residential gateway (RGWs), LAN proxy or base station (BS) which act under the supervision of an ISP.	Trace Driven Simulation
[106]	2008	VOD	Inter-ISP Traffic	To limit P2P traffic within the same ISP region, authors proposed an ISP-friendly rate allocation method.	Analytical Model and Simulation Study
[123]	2013	VOD	Inter-ISP Traffic	Proposed analytical model to analyze the relationship between CDN bandwidth and traffic localization at different levels inside the ISPs.	Simulation

6.2. Inter-ISP Traffic

Historically peer-to-peer systems have been ISP-oblivious and could generate significant amounts of the cross-ISP traffic - a fact which reportedly polarized ISPs attitude towards peer-to-peer systems [88]. While the peers within their local ISPs are able and are allowed to share their upstream bandwidth resources, the situation gets more complicated when the data traffic goes beyond the ISP boundaries for what the latter can be charged.

A centralized architecture to minimize the inter-ISP traffic has been proposed in [67]. The approach relies

on the Super Nodes (SN) deployed within each ISP and acting as central coordination units between the groups of peers and the CDN. In contrast, the authors of [78] proposed to employ the existing ISP-aware P2P protocols for PA-CDNs as a solution to handle the tussle between the ISPs and P2P networks. Thus, an ISP-aware PA-CDN in small ISPs can be built on top of a Tracker system, whereas DHT-based indexing such as Kadmelia can be more suitable for large ISPs. The authors also suggest the ALTO (application layer traffic optimization) protocol for the location aware peer selection. A similar solution for live streaming has been presented

Table 5: Architecture of the Commercial PA-CDNs

	Commercial PA-CDNs				
	Kankan [122]	**LiveSky** [61] [63] [117]	**NetSession** [124]	**Spotify** [35] [56] [120]	**Tudou** [64]
Service Type	VOD	Live	Mainly used for large size files such as software,updates but, also supports video streaming	On-Demand Music Streaming Service	User Generated Web Content
Number of Servers	258 edge servers, 18 Tracker servers, 24 control and management servers	400 edge cache server, 50 core service node (responsible for live streaming)	—	—	—
Number of P2P Users	30M	—	More than 25M users	10M	50M
Integration Architecture	Centralized PA-CDN Architecture	Centralized PA-CDN Architecture	Centralized PA-CDN Architecture	Centralized PA-CDN Architecture	Centralized PA-CDN Architecture
Request Redirection Mechanism	Server Redirection Mechanism	DNS Based Resolution	DNS Based Resolution	DNS Based Resolution	—

Summary of commercial PA-CDNs indicating *number of P2P users*, *number of servers*, type of *service* (VOD or Live), *integration architecture* and *request redirection mechanism*.

in [9]. To avoid the Inter-ISPs traffic a *proxy CDN node*, powered by a BitTorent-based streaming platform called NextShare, must be deployed either within the ISP network, or within a residential gateway (RGWs), LAN proxy or a base station (BS) under the supervision of the ISP. Such an intermediate layer of proxy nodes allows peers to indirectly connect between each other without a need of installing any additional client application, therefore, making it transparent for the users but also reducing peers' responsibilities to handle security, QoS, etc. Similarly, in [78] the authors rely on the Kademlia and ALTO protocols for peer tracking and matching.

Wang et al. [106] proposed an ISP-friendly rate allocation algorithm for peer-assisted VOD systems, in which the Tracker server selects the neighbours based on two key criteria: i) the peers belonging to the same autonomous system (AS) ; and ii) the peers arrived in the system at the same time. While the former criteria helps in localizing the traffic within the ISPs, the latter helps to optimize the stability of the peer-to-peer links.

A theoretical approach to understanding the relationship between CDN bandwidth and traffic localization at different levels inside the ISPs has been proposed in [123]. Particularly, the authors analyzed how different caching techniques can cooperate to reduce CDN load but also minimize the inter-ISP traffic. Their re-

sults suggest that a simple LRU (Least recently used) peers' cache replacement strategy is very effective for peer-assisted content delivery, whereas an increase in the inter-ISP traffic is shown to cause a linear reduction in the load on CDN.

7. Commercial Peer-Assisted CDNs

Various measurement studies have been conducted over the last decade to investigate the benefits and challenges of PA-CDNs in commercial settings. In this section, first we provide an introduction of five different commercial PA-CDNs. The review of this effort is summarized in Table 5. Then we discuss how various identified issues has been tackled by different operators. Table 6 provide summary of our analysis.

7.1. General Introduction and Integration Architecture

7.1.1. General Introduction

Xunlei is the 10th largest Internet company in China and **Kankan** is its peer-assisted on-demand streaming service with 31.4 million unique daily users as of the end of 2012 [122]. The Kankan's servers are organized into three geographic clusters around Beijing, Shanghai and Guangzhou and cover in total 19 metropolitan areas which together constitute around 5.4% of all Chinese cities. This setup captures 61.8% of subscribers

15

within the close proximity of CDN nodes. Kankan has deployed around 300 servers within the clusters out of which 258 are edge servers, 18 Tracker servers and few dozen remaining ones serving as control and management servers. As reported in [122] - 30M Kankan users have installed Xunlei accelerator to assist content distribution.

LiveSky was developed by China's leading content distribution provider - ChinaCache- in 2007. While ChinaCache provides content delivery services for both VOD and Live streaming applications, its peer assisted approach is specifically designed for live streaming. In 2007 ChinaCache deployed about 500 cache servers in 8 districts of China, out of which 50 are the core service nodes responsible for live streaming and over 400 edge caches deployed in a close proximity to users [117].

NetSession [124] is a global peer-assisted content distribution network, originally developed by Red Swoosh, a P2P content delivery company founded by Travis Kalanick and Michael Todd in 2001 and acquired by Akamai Technology in 2007 [86]. NetSession provides content delivery solutions to more than 25M users, mainly for big corporations and serves a user base from 299 countries. While most of Akamai NetSession's traffic constitutes large file distribution for software updates and games, it also serves multimedia traffic.

Spotify is a popular on-demand music streaming service which, according to [35] exploits peer-assistance to serve its 10 million-large user base around the world. Spotify is available for laptop clients as well as for several smart-phone platforms, however smart-phone clients do not participate in peer-assisted content distribution.

Tudou is one of the most popular user generated content websites in China, which uses a Tudou client-side plug-in installed by 50M of its users to assist content distribution. Interestingly, the average content length on Tudou is longer than that of Youtube and - although Tudou has a lower number of daily views - it surpasses Youtube in monthly traffic, i.e., 15 billion minutes of videos are served by Tudou per month versus 3 billion served by Youtube [64].

7.1.2. Integration Architecture

All of the commercial PA-CDNs we consider in this survey - Kankan, LiveSky, NetSession, Spotify and Tudou - are built on a pure two-tier centralized architecture in which multimedia files are split into small chunks be-
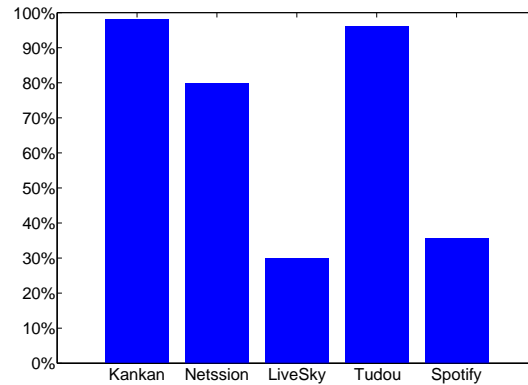

Figure 2: Potential traffic savings in commercial PA-CDNs from peer assistance in terms of total percentage of traffic offloaded from CDN to P2P.

fore they are delivered to peers. Peers discover their neighbors through a Tracker Server located within the CDN and download initial chunks of the content items as well as the content unavailable among peers from the closest edge servers.

7.1.3. Request Redirection Mechanism

NetSession and LiveSky [61] [63] use standard DNS request routing techniques and redirect users based on their location and the current load on the edge servers: if the nearest edge server is overloaded, the request is redirected to the next nearest and less loaded one. In contrast, Kankan uses the Server Redirection mechanism for load balancing. It handles flash crowds by dispatching requests among multiple close-by edge servers in parallel. A measurement study in [122] suggests that 45% and 20% of users' requests in Kankan are forwarded to the first nearest and the second nearest edge servers, respectively, when a flash crowd hits the system, whereas the rest of the requests are distributed equally among other nearby and less burden edge servers. In addition Kankan follows the three phase streaming mechanism proposed in [112], i.e., pure CDN streaming, hybrid CDN-P2P streaming and pure P2P streaming phase. With this setup the peer-assisted system shifts from pure CDN streaming to pure P2P streaming when the size of swarms becomes sufficiently large.

Similar to NetSession, Spotify's peer-assisted system also exploits the DNS query system to redirect users' requests to the nearest edge servers. However, in order to avoid single point of failure in case a Tracker server goes down, Spotify has adopted a dual peer selection mechanism. In addition to sending a request for the

16

Table 6: Detailed Comparison of the Commercial PA-CDNs

	Commercial PA-CDN				
	Kankan [122]	**LiveSky** [117] [63] [61]	**NetSession** [124]	**Spotify** [35] [56] [120]	**Tudou** [64]
QoS	Two streaming servers are assigned to clients for downloading initial chunks of a video content for fast startup.	Clients' buffer size has been reduced to half of the buffer size in pure P2P system for fast playback.	Edge server remains active during content downloading process and users download content in parallel from the edge server and peers.	To fulfill the initial request of the user and reduce music playback delay, the edge server forwards 15 seconds of music and at the same time searches for a set of peers that can serve the track.	—
In-Accessibility	—	STUN and UDP Protocol	STUN and UDP Protocol	TCP and Plug-n-play (UPnP) Protocol	—
Cache Management	Top 20% popular videos are cached at a CDN node and P2P system. Less popular videos are cached only at a CDN node.	—	No predictive caching policy is used. Content is cached when requested by client for a certain period of time to share with neighboring peers.	Reserve at most 10% of users' storage space and employs Least Recently Used (LRU) cache replacement policy for cache eviction.	A cache of 2GB is allocated on a user's device and content item replaced in a FIFO fashion.
Partial Participation and Incentives	—	High quality video streaming service for peers which download the P2P client software.	No incentive policy to encourage users' participation. By default peers are able to download content items in parallel from edge servers and neighboring peers.	—	—
Inter-ISP Traffic	Peers are localized within ISP regions.	Peers are localized within ISP regions.	Peers are localized within ISP regions.	—	P2P traffic is localized.

list of neighboring peers to the Tracker server, a peer also sends a request to all of neighboring peers in the P2P overlay which in their turn disseminate the request with their neighbors until an available uploading peer is found.

7.1.4. Traffic Savings in Commercial PA-CDNs

Different measurement studies have investigated and reported immense benefits from peer assistance in commercially deployed PA-CDNs (see Fig. 2 for summary). Ge Zhang et al. in [122] reported that, peer-assistance enables significant benefits for Kankan in distributing popular videos, i.e., up to 98.0% of the video content in Kankan are distributed in a peer-to-peer fashion, whereas edge nodes are responsible for handling a long-tail of unpopular videos. The authors in [61] [63] [117] built on observation that - despite the fact that the vast majority of users (90%) access the service from behind the firewalls - 30% of traffic can still be distributed via peer-assistance. Authors in [124] reported that - although only 31% of users in Akamai Netssion opt for sharing their upload bandwidth resources - 70%-80% of traffic can be offloaded to peers without any negative impact on the performance. The measurement results from [35] [56] [120] have reported that, only

8.8% of Spotify's music traffic is delivered from Spotify's servers, whereas the remaining 91.2% are delivered from either peer-assistance (35.8%) or from local caches (55.4%). Similarly, the authors in [64] have analyzed the user trace date collected from the Tudou platform and reported a traffic saving from 36% to 96% for popular video content.

7.2. Practical issues in commercial PA-CDNs

Next we discuss various practical issues faced in commercial deployments and the solutions adopted by different operators.

7.2.1. Quality of Service

A common approach to avoid startup delays lies in initializing video streams from the edge servers before switching to P2P overlay for streaming the rest of the video. For instance, Kankan employs a dual streaming strategy in which two types of servers, namely, Primary Content Servers (PCS) and Secondary Content Servers (SCS) are assigned to requesting peers to provide initial video chunks. SCSs store only initial chunks of the videos and serve users during the initial streaming phase, whereas the rest of the time users are served by PCSs which contains the full copy of the video.

As far as peer churn is concerned, LiveSky always maintains at least one connection to the edge server to guarantee backup video streaming if uploading peers suddenly disappear from the network. Spotify employs a similar strategy and always maintains a single TCP connection between a pair of clients and the server. To fulfill the initial request of a user and reduce the playback delay, the edge server forwards the first 15 seconds of an audio file while, at the same time, searching for the set of available peers which can serve the track. This is different from NetSession, in which parallel connections are maintained with both - an edge server and the list of peers in P2P system - and the content is downloaded simultaneously from both links, no matter how many peers are there to participate.

7.2.2. Inaccessibility

It is estimated that, 90% of users in China are located behind the firewalls [45] [61]. Among those, as the user trace collected [61] from LiveSky is revealed, 37% are not available for peer-to-peer distribution and nearly 30% can provide uploading capacity sufficient to serve only one user at a time. A measurement study conducted for Spotify has similarly reported that, only in 35% of cases peers were reachable for re-distributing the content. Overall, these results suggest that inaccessibility of peers behind the middle boxes is a major issue for peer-assisted content distribution.

To tackle the problem of peers' inaccessibility behind the middle boxes, LiveSky and NetSession have exploited the modified version of STUN [92] and UDP protocols. UDP protocol enables peers to communicate behind middle boxes, however, it does not provide a means to guarantee successful delivery. Therefore, LiveSky has adopted an application layer ACK mechanism to ensure reliable data delivery with UDP transmissions. It is worth mentioning that peers' resources are also utilized to facilitate the NAT traversing mechanism.

In contrast to LiveSky and NetSession, Spotify uses TCP and Universal Plug-n-Play (UPnP) protocols to cross the middle boxes. The UPnP protocol allows peers to accept incoming connection by opening port of a home router. When a peer wishes to establish a connection behind the firewall or NAT, it will attempt to establish a TCP connection with another peer which, in his turn, tries to establish a reciprocal connection. The peer-to-peer sharing is possible if both attempts to succeed.

7.2.3. Content Distribution/Cache Management

Some systems exploit pre-fetching to effectively distribute popular content in advance. Kankan, for instance, pre-fetches the top-20% popular videos to edge servers and peers, whereas unpopular videos are only cached on edge servers. Videos are also cached by peers while watching and might be deleted manually by the user or replaced by the cache management system after a certain period of time. In contrast, NetSession does not use predictive caching, however, it still allows peers to cache the content they watch for a certain amount of time and re-share it with other peers. Similarly, Tudou allocates a cache of 2GB on users' devices and replaces the content in a FIFO manner when the limit is reached. Similarly, Spotify reserves at most 10% of a user's storage space and employs the least recently used (LRU) cache replacement policy for cache eviction. Users can also manually configure the size of the cache up to 1GB [56]. In LiveSky a live stream is piped through an optimized route from the fixed, centralized backbone architecture to edge servers. Upon reaching edge servers, the stream is further divided into chunks and transferred through the layered P2P tree in each individual client region.

7.2.4. Copyright and Authentication

NetSession, Kankan and Tudou use HTTP(S) protocol to download video content from the edge servers which support some basic authentication, content integration and peer authorization functions. In multimedia streaming a server splits the content into segments before downloading and so the content might potentially be corrupted while transferring from one peer to another. There might be the cases when peers store different versions of the same file. To ensure integrity of video content, edge servers assign distinct IDs and a hash to each file to make it differentiable from other versions. These fields are used by a receiving peer who can validate the chunks of the content it receives, and discards corrupted ones. Similarly, when a peer joins the system, it registers in a system over and receives an authorization to distribute content to other peers directed by the Tracker server.

7.2.5. Partial Participation and Incentives

NetSession offers no direct incentives to discourage freeloading among peers. However, if a user enables upload bandwidth sharing option, he/she may receive incentives indirectly in the form of a faster downloading speed by receiving content in parallel from edge servers and peers simultaneously. Yet, it is reported in [124]

18

that only around 30% of NetSession users agree to participate in peer-assistance. LiveSky manages fairness among participating peers by offering high quality video streaming services to peers who download P2P client software, as compared to peers who are not willing to share their upload bandwidth. Overall, low participation level appears as a serious obstacle for commercial implementation of peer-assisted CDNs.

7.2.6. Inter ISP Traffic

It is worth noting, that all aforementioned commercial system employ some sort of location aware peer selection strategies. On one hand, peers are directed to the nearest edge servers using request redirection mechanism and, on the other hand, Tracker servers provide lists of only those peers that are in close proximity to the end-users and lie within the same region. This mechanism ensures that clients establish connections only with the peers who reside within the same ISP region. Similarly, in LiveSky each peer is localized within its edge server and peers are restricted to upload their resources to peers within the same region.

8. Conclusion and Vision on Direction of Research in the Future

In this paper, we present a comprehensive survey of the peer-assisted content delivery networks (PA-CDNs). We investigated the concept of merging two complementary technologies, key benefits in terms of the traffic saving, challenges and future research directions. Our analysis captures a decade-long history of research and industrial efforts in the field, including three feasibility studies for the large video-on-demand platforms (BBC iPlayer, MSN Video, Conviva), a review of five different commercial PA-CDNs (Kankan, LiveSky, Akamai NetSession, Spotify and Tudou) and a vast scope of technical and modeling papers. From the analysis of the literature, we identify three groups of challenging factors for industrial implementation of peer-assisted CDNs, namely, heterogeneity and scale challenges, technical challenges and commercial feasibility challenges, which form a taxonomy for our survey.

The results of our survey suggest a significant potential of peer-assisted content delivery in reducing infrastructural costs for content providers and CDNs: from around 50% to 88% of all consumer traffic can be offloaded to peers as reported by various feasibility studies. Moreover, our analysis of the five large commercial systems suggests that many technical and feasibility concerns with respect to deployment of PA-CDNs in production systems have been already addressed by the

existing peer-assisted CDN solutions. Among others, significant research efforts have been focused on solving quality of service issues, where bootstrapping video streams from edge servers and an effective management of P2P swarms have been proposed to solve the startup and playback latencies, respectively. As far as various swarm management strategies are concerned, our analysis suggests that all commercial PA-CDNs have implemented centralized architectures in which the P2P overlay is fully managed by the designated edge servers. Additionally, we identified the problem of inaccessibility of peers behind firewalls and middle boxes as an active area of research where many approaches, e.g., STUN, TURN, etc. have been proposed, but many obstacle factors, e.g., absence of a feedback mechanism, are yet to be solved.

Our analysis also revealed some of commercial feasibility issues which are yet to be addressed. This includes, among others a need for economically sound mechanisms to incentivize user participation, which is reportedly low in some of the existing systems [124] and is recognized as a significant obstacle factor in the others [38] [50]. The interests of ISPs are to be respected, too. One of the main political challenges for PA-CDN as recognized by numerous research papers have been so far in avoiding inter-ISP traffic known as a serious concern for the ISP business. It has been shown, in various recent papers, that by employing efficient topology-aware and ISP-friendly peer selection policies [67] [123] the deleterious influences of inter-ISP traffic can be minimized. Indeed, a study on the two leading CDN operators, i.e. Limelight and Akamai, demonstrated that, significant traffic savings (i.e., up to 66.53% for Akamai and 65.55% for Limlight) can be achieved, even if peer-to-peer traffic is localized within the ISP domains [44]. Similarly, a feasibility study for BBC iPlayer has reported potential traffic savings of up to 88% with ISP-friendly peer-assisted content delivery. However, [59] has raised an important concern with respect to ISP-friendly peer-assisted design, suggesting that localizing ISP traffic may negatively impact the quality of service. Therefore, a detailed investigation in this regard is required in future works.

Some of the commercial studies have recognized that effective strategies for managing end-user caches can significantly improve performance of peer-assisted content delivery networks, suggesting that the design of advanced content caching strategies - such as proactive caching and content pre-fetching - is a promising direction for future research.

Finally, the conventional paradigm of using P2P-assisted video content delivery through the Internet is

being shifted towards exploiting M2M and D2D communication and the widely exploited Internet of Things for visual data transmission and video content delivery. The fast expansion of Internet-of-Things (IoT) networks and the increasing demand for pervasive video have highlighted the need to support visual communications and video content delivery over machine-to-machine (M2M) networks [6] [95] [113]. Achieving this at appropriate quality and in an energy and delay-efficient manner is a major challenge for future IoT deployments.

The technologies and standards for IoT have recently advanced considerably, and paved the way for transmission of multimedia data over M2M networks. Specifically, Release-13 of 3GPP standards extends coverage and reduces cost, bandwidth and transmit power requirements for the so-called enhanced machine type communication (eMTC), a term used by 3GPP to refer to "cellular M2M" communications. With Cat-M1, a device category defined in Release-13, data rates of up to 1 Mbps (uplink and downlink) can be supported by eMTC over 1.4 MHz bandwidth, at a reduced peak transmit power (20 dBm). The first field trial of eMTC took place in early 2016 to validate its potential for addressing medium data rate IoT connectivity use cases [87], and standardization of a further lower-complexity device category (NB-IoT) that supports data rates of 200kbps completed in June '16. With respect to 'capillary M2M' networks, IEEE 802.15.4m task group has standardised full function (FFD) & reduced function device (RFD) categories, with full/reduced MAC capabilities, for cognitive M2M networks.

Two of the main emerging applications for M2M/D2D-assisted visual content delivery are:

1. **Transport of sensor visual data (or aggregates of such data) to cloud storage and edge servers for performing further analysis:**

   A key application of the emerging IoT is high-speed/low-power visual sensing and processing for surveillance, monitoring (e.g., vehicles, drones, visual similarity or classification for machine learning, visual search within IoT, medical and assisted living contexts). The main challenge in this category of applications is that the bandwidth and energy requirements of conventional high frame-rate video cameras are usually too high to be delivered through M2M/D2D networks. New paradigms are now emerging that challenge the very notion of video frames. In fact, biological vision does not use frames. Instead, it is based on detecting changes of scene illumination and movement in an asynchronous manner, while the visual cortex is "filling in" the remaining information. Inspired by this observation, hardware designs of neuromorphic sensors, a.k.a., dynamic vision sensors (DVS) or silicon retinas, have been proposed recently. DVS can enable transmission of visual data through M2M networks through a 10-fold reduction in power requirements (10-20 mW instead of 100s of mW of conventional video cameras) and up to a 100-fold increase in speed (e.g., DVS events can be rendered up to 2000 frames-per-sec.).

On the other hand, optimisation of uplink M2M networks has been studied recently, and energy-efficient machine-to-data aggregator communication has been investigated via PHY-layer optimization [71]. The combination of video coding and fountain coding for live video streaming over wireless networks has been proposed for IoT applications recently [104]. Finally, MAC-layer, network (NET)-layer, and application (APP) aware PHY-layer protocols suitable for M2M communications have been considered recently [3] [6] [87] [95].

2. **Improved quality and low delay video streaming through wireless caching and D2D communication:**

Wireless caching and M2M/D2D communication for video content delivery can be seen as an organic extension of the traditional P2P-assisted CDNs. D2D technology enables direct communication between devices located in close proximity without requiring to go through a central base station (BS). Some recent studies have shown that, cache-enabled Device-to-Device (D2D) communication can be helpful in improving spectral efficiency and reducing communication delay [36] [77] [107]. Mobile-CDNs are also gaining significant attention of the research community and mobile network operators (MNOs). Now MNOs have started to deploy their own CDN infrastructure for better control and management on the resources and bulk of video traffic [1] [2] [118]. Similar to peer-assisted-CDN, D2D-assisted Mobile-CDN can deliver significant benefits and a promising direction for future research.

References

[1] Jet stream cutting edge digital media logistics. Available from `http://www.akamai.com/uk/en/resources/mobile-cdn.jsp`.

[2] Mobile cdn. Available from `http://www.akamai.com/uk/en/resources/mobile-cdn.jsp`.

[3] www.3gpp.org. Available from `http://www.3gpp.org/ftp/Information/WORK_PLAN/Description_Releases/`.

[4] Limlight networks. Available from `http://https://www.limelight.com/delivery/`, Feb 2016.

[5] O. Abboud, A. Kovacevic, K. Graffi, K. Pussep, and R. Steinmetz. Underlay awareness in p2p systems: Techniques and challenges. In *Parallel & Distributed Processing, 2009. IPDPS 2009. IEEE International Symposium on*, pages 1–8. IEEE, 2009.

[6] A. Aijaz and A. H. Aghvami. Cognitive machine-to-machine communications for internet-of-things: a protocol stack perspective. *IEEE Internet of Things Journal*, 2(2):103–112, 2015.

[7] S. Androutsellis-Theotokis and D. Spinellis. A survey of peer-to-peer content distribution technologies. *ACM Computing Surveys (CSUR)*, 36(4):335–371, 2004.

[8] D. Applegate, A. Archer, V. Gopalakrishnan, S. Lee, and K. K. Ramakrishnan. Optimal content placement for a large-scale vod system. In *Proceedings of the 6th International COnference*, page 4. ACM, 2010.

[9] E. Baccaglini, M. Grangetto, E. Quacchio, and S. Zezza. A study of an hybrid cdn–p2p system over the planetlab network. *Signal Processing: Image Communication*, 27(5):430–437, 2012.

[10] F. Baccelli, F. Mathieu, I. Norros, and R. Varloot. Can p2p networks be super-scalable? In *INFOCOM, 2013 Proceedings IEEE*, pages 1753–1761. IEEE, 2013.

[11] A. Balachandran, V. Sekar, A. Akella, and S. Seshan. Analyzing the potential benefits of cdn augmentation strategies for internet video workloads. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 43–56. ACM, 2013.

[12] S. A. Baset and H. Schulzrinne. An analysis of the skype peer-to-peer internet telephony protocol. *arXiv preprint cs/0412017*, 2004.

[13] R. Bhagwan, S. Savage, and G. M. Voelker. Understanding availability. In *Peer-to-Peer Systems II*, pages 256–267. Springer, 2003.

[14] H. Bobarshad, M. Van der Schaar, A. H. Aghvami, R. S. Dilmaghani, and M. R. Shikh-Bahaei. Analytical modeling for delay-sensitive video over wlan. *Multimedia, IEEE Transactions on*, 14(2):401–414, 2012.

[15] H. Bobarshad, M. van der Schaar, and M. R. Shikh-Bahaei. A low-complexity analytical modeling for cross-layer adaptive error protection in video over wlan. *IEEE Transactions on Multimedia*, 12(5):427–438, 2010.

[16] V. BONNEAU. Telco cdn. Available from `http://www.idate.org/en/News/Telco-CDN_776.html`, 2015.

[17] M. Borella, J. Lo, D. Grabelsky, and G. Montenegro. Realm specific ip: framework. Technical report, 2001.

[18] R. Buyya, A.-M. Pathan, J. Broberg, and Z. Tari. A case for peering of content delivery networks. *Distributed Systems Online, IEEE*, 7(10):3–3, 2006.

[19] M. Calder, X. Fan, Z. Hu, E. Katz-Bassett, J. Heidemann, and R. Govindan. Mapping the expansion of google's serving infrastructure. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 313–326. ACM, 2013.

[20] L. Chen, Y. Zhou, M. Jing, and R. T. 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*, pages 43–48. ACM, 2015.

[21] Y.-C. Chen and W.-K. Jia. Challenge and solutions of nat traversal for ubiquitous and pervasive applications on the internet. *Journal of Systems and Software*, 82(10):1620–1626, 2009.

[22] X. Cheng and J. Liu. Nettube: Exploring social networks for peer-to-peer short video sharing. In *INFOCOM 2009, IEEE*, pages 1152–1160. IEEE, 2009.

[23] J.-w. Cho and Y. Yi. On the payoff mechanisms in peer-assisted services with multiple content providers: rationality and fairness. *Networking, IEEE/ACM Transactions on*, 22(3):731–744, 2014.

[24] Cisco. Cisco visual networking index: Global mobile data traffic forecast update 2014-2019 white paper. Available from `http://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white_paper_c11-520862.html`, Feb 2015.

[25] X. Cong, K. Shuang, S. Su, and F. Yang. An efficient server bandwidth costs decreased mechanism towards mobile devices in cloud-assisted p2p-vod system. *Peer-to-Peer Networking and Applications*, 7(2):175–187, 2014.

[26] H. Deng and J. Xu. Corepeer: A p2p mechanism for hybrid cdn-p2p architecture. In *Web-Age Information Management*, pages 278–286. Springer, 2013.

[27] M. El Dick, E. Pacitti, and B. Kemme. Flower-cdn: a hybrid p2p overlay for efficient query processing in cdn. In *Proceedings of the 12th International Conference on Extending Database Technology: Advances in Database Technology*, pages 427–438. ACM, 2009.

[28] A. Ferragut and F. Paganini. Fluid models of population and download progress in p2p networks. *IEEE Transactions on Control of Network Systems*, 3(1):34–45, 2016.

[29] B. Ford, P. Srisuresh, and D. Kegel. Peer-to-peer communication across network address translators. In *USENIX Annual Technical Conference, General Track*, pages 179–192, 2005.

[30] B. Frank, I. Poese, Y. Lin, G. Smaragdakis, A. Feldmann, B. Maggs, J. Rake, S. Uhlig, and R. Weber. Pushing cdn-isp collaboration to the limit. *ACM SIGCOMM Computer Communication Review*, 43(3):34–44, 2013.

[31] X. Fu, M. Stiemerling, and H. Schulzrinne. Implications and control of middleboxes in the internet. *Network, IEEE*, 22(5):6–7, 2008.

[32] G. Gao, Y. Wen, W. Zhang, and H. Hu. Cost-efficient and qos-aware content management in media cloud: Implementation and evaluation. In *Communications (ICC), 2015 IEEE International Conference on*, pages 6880–6886. IEEE, 2015.

[33] G. Gao, W. Zhang, Y. Wen, Z. Wang, and W. Zhu. Towards cost-efficient video transcoding in media cloud: Insights learned from user viewing patterns. *Multimedia, IEEE Transactions on*, 17(8):1286–1296, 2015.

[34] M. Garmehi and M. Analoui. Envy-free resource allocation and request routing in hybrid cdn–p2p networks. *Journal of Network and Systems Management*, pages 1–32.

[35] M. Goldmann and G. Kreitz. Measurements on the spotify peer-assisted music-on-demand streaming system. In *Peer-to-Peer Computing (P2P), 2011 IEEE International Conference on*, pages 206–211. IEEE, 2011.

[36] 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 Transac-*

tions on Wireless Communications, 13(7):3665–3676, 2014.

[37] B. Gowan. What netflix's 1 billion video hours streamed means for bandwidth. Available from http://www.ciena.com/connect/blog/What-Netflixs-1-billion-video-hours-streamed- means-for-bandwidth.html, Aug 2012.

[38] S. Gramatikov and F. Jaureguizar. Modelling and analysis of non-cooperative peer-assisted vod streaming in managed networks. *Multimedia Tools and Applications*, 75(8):4321–4348, 2016.

[39] D. H. Ha, T. Silverton, and O. Fourmaux. *A novel hybrid CDN-P2P mechanism for effective real-time media streaming*. PhD thesis, M. Sc. Thesis, 2008.

[40] X. Hei, C. Liang, J. Liang, Y. Liu, and K. W. Ross. A measurement study of a large-scale p2p iptv system. *Multimedia, IEEE Transactions on*, 9(8):1672–1687, 2007.

[41] C. Hu, M. Chen, C. Xing, and B. Xu. Eue principle of resource scheduling for live streaming systems underlying cdn-p2p hybrid architecture. *Peer-to-Peer Networking and Applications*, 5(4):312–322, 2012.

[42] C. Huang, J. Li, and K. W. Ross. Can internet video-on-demand be profitable? *ACM SIGCOMM Computer Communication Review*, 37(4):133–144, 2007.

[43] C. Huang, A. Wang, J. Li, and K. W. Ross. Measuring and evaluating large-scale cdns. In *ACM IMC*, volume 8, 2008.

[44] C. Huang, A. Wang, J. Li, and K. W. Ross. Understanding hybrid cdn-p2p: why limelight needs its own red swoosh. In *Proceedings of the 18th International Workshop on Network and Operating Systems Support for Digital Audio and Video*, pages 75–80. ACM, 2008.

[45] S. Hyunjoong. Pcubecast: A novel peer-assisted live streaming system. In *Peer-to-Peer Computing (P2P), 2011 IEEE International Conference on*, pages 212–215. IEEE, 2011.

[46] S. Ioannidis and P. Marbach. On the design of hybrid peer-to-peer systems. In *ACM SIGMETRICS Performance Evaluation Review*, volume 36, pages 157–168. ACM, 2008.

[47] Y. Jin, G. Kesidis, J. Shin, F. Kocak, and Y. Yi. Impacts of selfish behaviors on the scalability of hybrid client–server and peer-to-peer caching systems.

[48] Y. Jin, Y. Wen, and K. Guan. Toward cost-efficient content placement in media cloud: Modeling and analysis. *IEEE Transactions on Multimedia*, 18(5):807–819, 2016.

[49] S. Kang and H. Yin. A hybrid cdn-p2p system for video-on-demand. In *Future Networks, 2010. ICFN'10. Second International Conference on*, pages 309–313. IEEE, 2010.

[50] D. Karamshuk, N. Sastry, A. Secker, and J. Chandaria. Isp-friendly peer-assisted on-demand streaming of long duration content in bbc iplayer. *in Proc. IEEE INFOCOM*, 2015.

[51] D. Karamshuk, N. Sastry, A. Secker, and J. Chandaria. On factors affecting the usage and adoption of a nation-wide tv streaming service. *in Proc. IEEE INFOCOM*, 2015.

[52] S. Kaune, R. C. Rumin, G. Tyson, A. Mauthe, C. Guerrero, and R. Steinmetz. Unraveling bittorrent's file unavailability: measurements and analysis. In *Peer-to-Peer Computing (P2P), 2010 IEEE Tenth International Conference on*, pages 1–9. IEEE, 2010.

[53] T. N. Kim, S. Jeon, and Y. Kim. A cdn-p2p hybrid architecture with content/location awareness for live streaming service networks. In *Consumer Electronics (ISCE), 2011 IEEE 15th International Symposium on*, pages 438–441. IEEE, 2011.

[54] Y. Kim, Y. Kim, H. Yoon, and I. Yeom. Peer-assisted multimedia delivery using periodic multicast. *Information Sciences*, 298:425–446, 2015.

[55] U. C. Kozat, Ö. Harmancı, S. Kanumuri, M. U. Demircin, and M. R. Civanlar. Peer assisted video streaming with supply-demand-based cache optimization. *Multimedia, IEEE Transactions on*, 11(3):494–508, 2009.

[56] G. Kreitz and F. Niemelä. Spotify–large scale, low latency, p2p music-on-demand streaming. In *Peer-to-Peer Computing (P2P), 2010 IEEE Tenth International Conference on*, pages 1–10. IEEE, 2010.

[57] Z. Li, J. Lin, M.-I. Akodjenou, G. Xie, M. A. Kaafar, Y. Jin, and G. Peng. Watching videos from everywhere: a study of the pptv mobile vod system. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 185–198. ACM, 2012.

[58] Z. Li, H. Shen, H. Wang, G. Liu, and J. Li. Socialtube: P2p-assisted video sharing in online social networks. In *INFOCOM, 2012 Proceedings IEEE*, pages 2886–2890. IEEE, 2012.

[59] M. Liu, F. Lu, X. Luo, and Z. Qin. An isp-friendly hierarchical overlay for p2p live streaming. In *Proceedings of 14-th IEEE International Conference on Peer-to-Peer Computing*, 2014.

[60] X. Liu, F. Dobrian, H. Milner, J. Jiang, V. Sekar, I. Stoica, and H. Zhang. A case for a coordinated internet video control plane. In *Proceedings of the ACM SIGCOMM 2012 conference on Applications, technologies, architectures, and protocols for computer communication*, pages 359–370. ACM, 2012.

[61] X. Liu, H. Yin, and C. Lin. A novel and high-quality measurement study of commercial cdn-p2p live streaming. In *Communications and Mobile Computing, 2009. CMC'09. WRI International Conference on*, volume 3, pages 325–329. IEEE, 2009.

[62] Y. Liu, Y. Guo, and C. Liang. A survey on peer-to-peer video streaming systems. *Peer-to-peer Networking and Applications*, 1(1):18–28, 2008.

[63] Y. Liu, H. Yin, G. Zhu, and X. Liu. Peer-assisted content delivery network for live streaming: Architecture and practice. In *Networking, Architecture, and Storage, 2008. NAS'08. International Conference on*, pages 149–150. IEEE, 2008.

[64] Z. Liu, Y. Ding, Y. Liu, and K. Ross. Peer-assisted distribution of user generated content. In *Peer-to-Peer Computing (P2P), 2012 IEEE 12th International Conference on*, pages 261–272. IEEE, 2012.

[65] Z. Lu, X. Gao, S. Huang, and Y. Huang. Scalable and reliable live streaming service through coordinating cdn and p2p. In *Parallel and Distributed Systems (ICPADS), 2011 IEEE 17th International Conference on*, pages 581–588. IEEE, 2011.

[66] Z. Lu, Y. Wang, and Y. R. Yang. An analysis and comparison of cdn-p2p-hybrid content delivery system and model. *journal of communications*, 7(3):232–245, 2012.

[67] Z. Lu, J. Wu, L. Chen, S. Huang, and Y. Huang. Cph-vod: A novel cdn–p2p-hybrid architecture based vod scheme. In *Web Information Systems Engineering–WISE 2010*, pages 578–586. Springer, 2010.

[68] E. K. Lua, J. Crowcroft, M. Pias, R. Sharma, and S. Lim. A survey and comparison of peer-to-peer overlay network schemes. *Communications Surveys & Tutorials, IEEE*, 7(2):72–93, 2005.

[69] M. Ma, Z. Wang, K. Su, and L. Sun. Understanding content placement strategies in smartrouter-based peer cdn for video streaming. *arXiv preprint arXiv:1605.07705*, 2016.

[70] R. Mahy, P. Matthews, and J. Rosenberg. Traversal using relays around nat (turn): Relay extensions to session traversal utilities for nat (stun). Technical report, 2010.

[71] D. Malak, H. S. Dhillon, and J. G. Andrews. Optimizing data aggregation for uplink machine-to-machine communication networks. *IEEE Transactions on Communications*, 64(3):1274–1290, 2016.

[72] E. Markakis, D. Negru, J. Bruneau-Queyreix, E. Pallis, G. Mastorakis, and C. X. Mavromoustakis. A p2p home-box

overlay for efficient content distribution. *Emerging Innovations in Wireless Networks and Broadband Technologies*, page 199, 2016.

[73] D. S. Menasche, A. A. Rocha, B. Li, D. Towsley, and A. Venkataramani. Content availability and bundling in swarming systems. In *Proceedings of the 5th international conference on Emerging networking experiments and technologies*, pages 121–132. ACM, 2009.

[74] V. Misra, S. Ioannidis, A. Chaintreau, and L. Massoulié. Incentivizing peer-assisted services: a fluid shapley value approach. In *ACM SIGMETRICS Performance Evaluation Review*, volume 38, pages 215–226. ACM, 2010.

[75] A. Mondal, I. Trestian, Z. Qin, and A. Kuzmanovic. P2p as a cdn: A new service model for file sharing. *Computer Networks*, 56(14):3233–3246, 2012.

[76] T. Mori, N. Kamiyama, S. Harada, H. Hasegawa, and R. Kawahara. Improving deployability of peer-assisted cdn platform with incentive. In *Global Telecommunications Conference, 2009. GLOBECOM 2009. IEEE*, pages 1–7. IEEE, 2009.

[77] M. Naslcheraghi, S. A. Ghorashi, and M. Shikh-Bahaei. Full-duplex device-to-device communication for wireless video distribution. *IET Communications*, Jan 2017.

[78] Y. Nicolas, D. Wolff, D. Rossi, and A. Finamore. I tube, youtube, p2ptube: Assessing isp benefits of peer-assisted caching of youtube content. In *Peer-to-Peer Computing (P2P), 2013 IEEE Thirteenth International Conference on*, pages 1–2. IEEE, 2013.

[79] A. Passarella. A survey on content-centric technologies for the current internet: Cdn and p2p solutions. *Computer Communications*, 35(1):1–32, 2012.

[80] A.-M. K. Pathan and R. Buyya. A taxonomy and survey of content delivery networks. *Grid Computing and Distributed Systems Laboratory, University of Melbourne, Technical Report*, 2007.

[81] R. Peterson and E. G. Sirer. Antfarm: Efficient content distribution with managed swarms. In *NSDI*, volume 9, pages 107–122, 2009.

[82] M. H. Pinson and S. Wolf. A new standardized method for objectively measuring video quality. *Broadcasting, IEEE Transactions on*, 50(3):312–322, 2004.

[83] D. Qiu and R. Srikant. Modeling and performance analysis of bittorrent-like peer-to-peer networks. In *ACM SIGCOMM Computer Communication Review*, volume 34, pages 367–378. ACM, 2004.

[84] S. Ratnasamy, P. Francis, M. Handley, R. Karp, and S. Shenker. *A scalable content-addressable network*, volume 31. ACM, 2001.

[85] S. Ratnasamy, M. Handley, R. Karp, and S. Shenker. Topologically-aware overlay construction and server selection. In *INFOCOM 2002. Twenty-First Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 3, pages 1190–1199. IEEE, 2002.

[86] Red. Redswoosh. Available from http://en.wikipedia.org/wiki/RedSwoosh.

[87] P. Release. Kt, nokia conduct world's first emtc field trial on lte network. Available from http://company.nokia.com/en/news/press-releases/2016/01/26/kt-nokia-conduct-worlds-first-emtc-field-trial-on-lte-network, Jan 2016.

[88] R. Rodrigues and P. Druschel. Peer-to-peer systems. *Communications of the ACM*, 53(10):72–82, 2010.

[89] P. Rodriguez, S.-M. Tan, and C. Gkantsidis. On the feasibility of commercial, legal p2p content distribution. *ACM SIGCOMM Computer Communication Review*, 36(1):75–78, 2006.

[90] J. Rosenberg. Interactive connectivity establishment (ice): A methodology for network address translator (nat) traversal for offer/answer protocols. 2010.

[91] J. Rosenberg, R. Mahy, C. Huitema, and J. Weinberger. Stun-simple traversal of udp through network address translators. 2003.

[92] J. Rosenberg, R. Mahy, P. Matthews, and D. Wing. Rfc 5389: Session traversal utilities for nat (stun). *Internet Engineering Task Force*, 2008.

[93] R. Roverso, A. Naiem, M. Reda, M. El-Beltagy, S. El-Ansary, N. Franzen, and S. Haridi. On the feasibility of centrally-coordinated peer-to-peer live streaming. In *Consumer Communications and Networking Conference (CCNC), 2011 IEEE*, pages 1061–1065. IEEE, 2011.

[94] J. Rückert, T. Knierim, and D. Hausheer. Clubbing with the peers: A measurement study of bittorrent live. In *14-th IEEE International Conference on Peer-to-Peer Computing*, pages 1–10. IEEE, 2014.

[95] H. Saki and M. Shikh-Bahaei. Cross-layer resource allocation for video streaming over ofdma cognitive radio networks. *Multimedia, IEEE Transactions on*, 17(3):333–345, 2015.

[96] S. Saroiu, K. P. Gummadi, R. J. Dunn, S. D. Gribble, and H. M. Levy. An analysis of internet content delivery systems. *ACM SIGOPS Operating Systems Review*, 36(SI):315–327, 2002.

[97] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft. Track globally, deliver locally: improving content delivery networks by tracking geographic social cascades. In *Proceedings of the 20th international conference on World wide web*, pages 457–466. ACM, 2011.

[98] H. Shen, Z. Li, Y. Lin, and J. Li. Socialtube: P2p-assisted video sharing in online social networks. *Parallel and Distributed Systems, IEEE Transactions on*, 25(9):2428–2440, 2014.

[99] P. Sijben. Middlebox communications (midcom) protocol requirements. *syntax and semantics*, 2:1, 2002.

[100] E. Siu. Ignore video and miss out on 69 percent of mobile traffic. Available from http://www.entrepreneur.com/article/237644, sep 2014.

[101] K.-A. Skevik, V. Goebel, and T. Plagemann. Analysis of bittorrent and its use for the design of a p2p based streaming protocol for a hybrid cdn. *Delft University of Technology Parallel and Distributed Systems Report Series, Tech. Rep. Technical Report*, 2004.

[102] K.-A. Skevik, V. Goebel, and T. Plagemann. Design of a hybrid cdn. In *Interactive Multimedia and Next Generation Networks*, pages 206–217. Springer, 2004.

[103] D. Stutzbach and R. Rejaie. Understanding churn in peer-to-peer networks. In *Proceedings of the 6th ACM SIGCOMM conference on Internet measurement*, pages 189–202. ACM, 2006.

[104] K. Sun, H. Zhang, D. Wu, and H. Zhuang. Mpc-based delay-aware fountain codes for real-time video communication. *IEEE Internet of Things Journal*.

[105] D. S. Wallach. A survey of peer-to-peer security issues. In *Software Security—Theories and Systems*, pages 42–57. Springer, 2003.

[106] J. Wang, C. Huang, and J. Li. On isp-friendly rate allocation for peer-assisted vod. In *Proceedings of the 16th ACM international conference on Multimedia*, pages 279–288. ACM, 2008.

[107] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. Leung. Cache in the air: exploiting content caching and delivery techniques for 5g systems. *IEEE Communications Magazine*, 52(2):131–139, 2014.

[108] Z. Wang, L. Sun, X. Chen, W. Zhu, J. Liu, M. Chen, and S. Yang. Propagation-based social-aware replication for social video contents. In *Proceedings of the 20th ACM international*

conference on Multimedia, pages 29–38. ACM, 2012.

[109] P. Wendell and M. J. Freedman. Going viral: flash crowds in an open cdn. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, pages 549–558. ACM, 2011.

[110] M. Wichtlhuber, S. Dargutev, S. Mullery, A. Kleiny, and D. Hausheer. Qtrade: a quality of experience based peercasting trading scheme. In *Peer-to-Peer Computing (P2P), 2015 IEEE International Conference on*, pages 1–10. IEEE, 2015.

[111] J. Wu, Z. Lu, B. Liu, and S. Zhang. Peercdn: A novel p2p network assisted streaming content delivery network scheme. In *Computer and Information Technology, 2008. CIT 2008. 8th IEEE International Conference on*, pages 601–606. IEEE, 2008.

[112] D. Xu, S. S. Kulkarni, C. Rosenberg, and H.-K. Chai. Analysis of a cdn–p2p hybrid architecture for cost-effective streaming media distribution. *Multimedia Systems*, 11(4):383–399, 2006.

[113] J. Xu, Y. Andrepoulos, Y. Xiao, and M. van Der Schaar. Non-stationary resource allocation policies for delay-constrained video streaming: Application to video over internet-of-things-enabled networks. *IEEE Journal on Selected Areas in Communications*, 32(4):782–794, 2014.

[114] Y. Xu, C. Yu, J. Li, and Y. Liu. Video telephony for end-consumers: measurement study of google+, ichat, and skype. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, pages 371–384. ACM, 2012.

[115] H.-C. Yang, M.-Y. Hsieh, H.-F. Yu, and L.-M. Tseng. A replication-aware cdn-p2p architecture based on two-step server selection and network coding. In *Advances in Multimedia Information Processing-PCM 2008*, pages 738–747. Springer, 2008.

[116] Z. Yao, D. Leonard, X. Wang, and D. Loguinov. Modeling heterogeneous user churn and local resilience of unstructured p2p networks. In *Network Protocols, 2006. ICNP'06. Proceedings of the 2006 14th IEEE International Conference on*, pages 32–41. IEEE, 2006.

[117] H. Yin, X. Liu, T. Zhan, V. Sekar, F. Qiu, C. Lin, H. Zhang, and B. Li. Livesky: Enhancing cdn with p2p. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 6(3):16, 2010.

[118] F. Z. Yousaf, M. Liebsch, A. Maeder, and S. Schmid. Mobile cdn enhancements for qoe-improved content delivery in mobile operator networks. *IEEE Network*, 27(2):14–21, 2013.

[119] H. Yu, D. Zheng, B. Y. Zhao, and W. Zheng. Understanding user behavior in large-scale video-on-demand systems. In *ACM SIGOPS Operating Systems Review*, volume 40, pages 333–344. ACM, 2006.

[120] B. Zhang, G. Kreitz, M. Isaksson, J. Ubillos, G. Urdaneta, J. Pouwelse, D. Epema, et al. Understanding user behavior in spotify. In *INFOCOM, 2013 Proceedings IEEE*, pages 220–224. IEEE, 2013.

[121] G. Zhang, W. Liu, X. Hei, and W. Cheng. Unreeling xunlei kankan: Understanding hybrid cdn-p2p video-on-demand streaming. 2013.

[122] G. Zhang, W. Liu, X. Hei, and W. Cheng. Unreeling xunlei kankan: understanding hybrid cdn-p2p video-on-demand streaming. *IEEE Transactions on Multimedia*, 17(2):229–242, 2015.

[123] J. Zhao, C. Wu, and X. Lin. Locality-aware streaming in hybrid p2p-cloud cdn systems. *Peer-to-Peer Networking and Applications*, 8(2):320–335, 2013.

[124] M. Zhao, P. Aditya, A. Chen, Y. Lin, A. Haeberlen, P. Druschel, B. Maggs, B. Wishon, and M. Ponec. Peer-assisted content distribution in akamai netsession. In *Proceedings of the 2013 conference on Internet measurement conference*, pages 31–42. ACM, 2013.

[125] M. Zink, K. Suh, Y. Gu, and J. Kurose. Characteristics of youtube network traffic at a campus network–measurements, models, and implications. *Computer Networks*, 53(4):501–514, 2009.

Biography

Nasreen Anjum

Nasreen Anjum is a PhD student at Department of Informatics King's College London. She received her MS degree in Computer Science from City University of Science and Information Technology Peshawar, in 2008. She is with the Shaheed Benazir Bhutto Women University Peshawar since 2007 as, lecturer in Computer Science. She is a recipient of fellowship award 2014, 2015 and 2016 from the "Schlumberger faculty for the future fellowship award". Her research interests include CDN, Mobile-CDN, P2P, device-to-device networks, and machine- to-machine communication.

Dr. Dmytro Karamshuk

Dmytro Karamshuk is a postdoctoral researcher at King's College London. His research interest is in machine learning approaches to modelling behavior of online users. He has previously worked on studying watching patterns of TV-streaming users (e.g., BBC iPlayer), mining geo-location social networks (e.g., Foursquare) and predicting user behaviour in image-based social platforms (e.g., Pinterest). He is an active contributor to the computer networks (Infocom, ComMag, etc.) and data mining communities (KDD, WWW, etc.).

Dr. Mohammad Shikh-Bahaei

Dr. Mohammad Shikh-Bahaei received the B.Sc. degree from the University of Tehran, Tehran, Iran, in 1992, the M.Sc. degree from the Sharif University of Technology, Tehran, Iran, in 1994, and the Ph.D. degree from King's College London, U.K., in 2000. He has worked for two start-up companies, and for National Semiconductor Corp., Santa Clara, CA, USA (now part of Texas Instruments Incorporated), on the design of third-generation (3G) mobile handsets, for which he has been awarded three U.S. patents as inventor and co-inventor, re-

spectively. In 2002, he joined King's College London as Lecturer, and is now a Reader with the Center for Telecommunications Research, Department of Informatics, King's College London. He has since authored or co-authored numerous journal and conference articles. He has been engaged in research in the area of wireless communications and signal processing for 20 years both in academic and industrial organizations. His research interests include the fields of resource allocation for multimedia applications over heterogeneous communication networks, and his recent research also encompasses applications in healthcare. Dr. Shikh-Bahaei was the Founder and Chair of the Wireless Advanced (for-merly SPWC) Annual International Conference from 2003 to 2012. He was the recipient of the overall King's College London Excellence in Supervisory Award in 2014.

Dr. Nishanth Sastry

Nishanth Sastry is a Senior Lecturer at King's College London. He holds a PhD from the University of Cambridge, UK, a Master's degree from The University of Texas at Austin, and a Bachelor's degree from Bangalore University, India, all in Computer Science. He has over six years of experience in the Industry (Cisco Systems, India and IBM Software Group, USA) and Industrial Research Labs (IBM TJ Watson Research Center).

25