CONCURRENCY AND COMPUTATION: PRACTICE AND EXPERIENCE
Concurrency Computat.: Pract. Exper. (2015)
Published online in Wiley Online Library (wileyonlinelibrary.com). DOI: 10.1002/spe.3464

SPECIAL ISSUE PAPER

Demystifying commercial content delivery networks in China

Hao Yin[1,*,†], Bo Qiao[1], Yan Luo[2], Chen Tian[3] and Y. Richard Yang[4]

[1]*Dept.t of CS, Tsinghua University, Beijing, China*
[2]*Dept. of ECE, University of Massachusetts Lowell, Lowell, MA, USA*
[3]*Dept. of EIE, Huazhong Univ. of Sci. and Tech., Wuhan, China*
[4]*Dept. of CS, Yale University, New Haven, CT, USA*

SUMMARY

Over the past decade, content delivery networks (CDNs) have attracted substantial Internet traffic and improved quality of experience for Internet users. However, the evolution of the Internet ecosystem, which is driven by underlying economic incentives and ever emerging technologies, posts great challenges to the existing commercial CDNs (CCDNs). Thoroughly understanding the CDN industry from different aspects including market choice, technology, performance, tendency and infrastructure is indispensable to future Internet. In this paper, we conduct the first comprehensive study of China's CDNs using continuous, at-scale, content-driven measurements. Based on the massive amount of measurement data with multidimensional properties, we demystify the CCDNs in China and answer two important questions: (1) what is the development trend of CCDNs in China and (2) what are their unique characteristics. The answers to these questions have significant implications on CDN providers and users. Copyright © 2015 John Wiley & Sons, Ltd.

Received 25 September 2014; Accepted 24 November 2014

KEY WORDS: CDN

1. INTRODUCTION

1.1. Motivation

To bridge the gap between network applications and best effort Internet provider (IP) networks, the technique of content delivery networks (CDNs) has emerged and been developed over the last decade. CDNs offer multimedia data, serve static content and support emerging applications. CDN service has evolved to a multibillion dollar business and functions as the vehicle for nearly half of the Internet backbone traffic as of 2012 [1].

Understanding the CDN industry from different aspects is indispensable to future Internet and network applications such as news portal, video website, social network system, e-commerce and so on. Here, we use Internet content provider (ICP) to broadly represent the owners of Internet applications. User experience is of paramount importance to ICPs, thus CDN providers constantly strive to improve QoS through infrastructure and control optimizations [2–6]. An emerging approach is to leverage a global view of CDN performance, service cost, deploy infrastructure and so on to dynamically direct users to suitable CDNs to improve user experience [7]. To facilitate such novel enhancement schemes, continuous, accurate and comprehensive measurements are crucial.

From content hosting perspective, ICPs need to make important decisions on how to place contents in CDNs to meet QoS requirements and minimize the cost. Past and current market choices

*Correspondence to: Hao Yin, Research Institute of Information Technology.
†E-mail: h-yin@mail.tsinghua.edu.cn

of CDNs together imply the technical advantages of one CDN over another. Thus, a comprehensive study on how top ICPs distribute their data over CDNs and the trend of such choice over the years can provide important insights and help make provisioning decisions.

Despite the large amount or measurement for North America [2, 8–17], the CDN industry in the second biggest Internet market, China, remains mystery to outsiders. CDNs in China have undergone a rapid growth in the last decade. Commercial CDN (CCDN) providers vary significantly in size, deployment and management. Besides leveraging CCDNs, an increasing number of ICPs have begun the deployment of their own private CDNs (also referred to as PCDNs in this paper) for control and performance optimization, similar to what is happening in North America [18]. Such an interesting trend shows that ICPs' choices of CDNs are neither static nor exclusive. An in-depth study of China's CDN state and trend, as well as its comparison with well-studied regions such as North America, are both timely and unique.

1.2. Challenges

Measuring CDNs' performance and characterizing their content data is very challenging for a number of reasons. First, there is no incentive for CDN providers to reveal such information to the public. The contractual information on content types and volume is not publicly accessible. Second, using a black-box approach to measure and compare multiple CDN providers requires an objective and fair framework that encompasses representative ICPs and CDN providers. One needs to derive an innovative approach to collect data without losing generality and impartiality. Third, multifaceted and continuous measurements on temporal, geographical and ISP dimensions demand tremendous resources on infrastructures and time. To establish such a measurement infrastructure and operate it for years are not trivial.

We are motivated to demystify China's CDN industry by addressing these challenges. In this paper, we utilize a content-driven active measurement infrastructure deployed at scale on China's Internet to quantify the content types and volume, delivery performance and infrastructure hierarchy of CDN providers. We study China's top 100 websites ranked by Alexa [19] and analyze over nine million web content links, using over 60 measurement nodes deployed in 40 cities throughout China's major ISPs. We conduct measurement experiments in both 2011 and 2013 with nearly identical approaches, allowing direct comparisons and observing the CDN trends in China.

With such a large scale continuous content driven measurement, we are able to obtain a massive amount of data with multidimensional properties: time, ICPs, CDNs, ISPs and geographic locations. Analysis on such a big dataset empowers us to reveal much hidden information of CCDNs in terms of the state and trend of the market share, infrastructure, performance and so on.

1.3. Contributions

Specifically, we make the following contributions in this paper.

- *Unveil market segmentation of CCDNs using content-driven measurement.* Our detailed analysis shows the market choice of CCDNs and PCDNs and reveals the rapid growth of PCDNs, which have reshaped the CDN industry in China. We find that CCDNs in China mainly carry static contents, while PCDNs transport revenue-generating contents (e.g. video).
- *Understand the trend of CCDN development with a multidimensional study.* Our at-scale measurement infrastructure has been operating for 2 years and the collected data draw a full picture of China's CDNs in multiple dimensions: temporal (year 2011 and 2013), geographical (40 provinces), ISPs (all major ISPs) and content-types (video, image and text). It is by far the most up-to-date and comprehensive study of China's CDN industry.
- *Uncover unique characteristics of China's CCDNs.* We show that the characteristics of CCDNs differ significantly from those in developed regions such as the North America: (a) the architecture of CCDNs is homogeneous in China (i.e. enter-deep-into-ISP) compared with the heterogeneous CDN architectures in the USA; and (b) China's ISPs do not play any significant roles in CDN market whereas a couple of American ISPs play major roles (e.g. Level 3 delivers 25% of CDN traffic [17]).

DOI: 10.1002/cpe

1.4. *Implications*

Based on the measurement results, we derive some implications, which have significant impacts on all the stakeholder of CDN industry.

- **CDN coverage in China** CDNs in China have gained more attention from ICPs. The usage of CDNs by top 100 ICPs increases significantly from 2011 to 2013. The number of ICPs having over half of their content volume carried by CDNs grows from 26% to 55%. Over 35% of the ICPs have 80% of their contents delivered with CDNs in 2013. While the market share is dominated by a few large CCDN providers, the coverage of CDN still has room to grow in terms of number of ICPs and content volumes.
- **Commercial versus PCDNs** A significant trend is that top ICPs have increasingly strong technical expertise and financial advantages, which enable the deployment of their own resources for flexible content control and optimization. CCDN providers are facing strong competitions from existing ICPs. Understanding the choices of CDNs made by current ICPs helps CDN providers improve the QoS and optimize their infrastructure.
- **Homogeneous architecture or heterogeneous one** While the degree of clustering and geographical distribution of CDN servers differ among providers, all CDNs in China use Akamai-like enters-into-ISPs approach with a two-tier DNS server infrastructure. The homogeneous architecture of China's CDNs is in sharp contrast to those in the USA, where CDNs follow various architectures (e.g. deep-into-ISP and bring-ISP-to-home). China's CCDNs should rethink the alternative architecture for lowering management costs and maximizing resource utilization on different ISP networks.

The paper is organized as follows. Section 2 discusses related work. We describe the measurement methodology and summary statistics in Section 3. The analysis of China's CDN industry status and trends is presented in Section 4. CDN charting is in Section 5. The QoS metrics of CDNs are compared in Section 6. Finally, Section 7 concludes the paper.

2. RELATED WORK

There have been many measurement studies on content distribution network [2, 8–17]. In this section, we briefly summarize those results and compare them with our measurement in China.

2.1. *Technology*

There are two CDN design philosophies [10]. The first is to *enter deep into ISPs*: by deploying surrogate servers inside ISPs' point of presence, the cached content is close to users. The leading CDN Akamai is of this type. Another philosophy is to *bring ISPs to home*, which builds large data centers and peering these centers extensively with ISPs. The well-known representative CDN of this type is Limelight. As shown in Section 5.1, all major CDN providers in China belong to the first type.

DNS redirection and *application request routing* are two commonly used technologies for directing user requests to servers in a particular CDN. Most CDN providers use DNS redirection, which resolve the requested canonical name (CNAME) to the IP address of a selected surrogate server [8, 9]. Application request routing was used only by a few small obsolete CDNs, such as Clearway and Fasttide [9]. Nowadays, all major players in China CDN market adopt DNS redirection (see Section 5.2 for details).

2.2. *Market*

The CDN market share is inherently dynamic. Krishnamurthy *et al.* [9] studied the usage of seven popular CDNs in North America circa 2001. After more than a decade, only Akamai is still active in this arena [17]. At that time, 31% of the top 127 (and 17% of the top 1030) sites adopted CCDN services. In 12-year's later China, 75 of the top 100 sites use CCDN services (Section 4.1).

DOI: 10.1002/cpe

Now, let's look at the market share comparison among CDNs. Regarding the number of customers, in 2001, 37 out of 39 (and 165 of 177) top sites used Akamai. With respect to the number of objects, Akamai also served 85–98% objects at that time. About the share of traffic volume, DeepField has a more recent result in 2012 [17]: Akamai, Level 3 and Limelight occupy 48%, 25% and 18% of the total CDN traffic, respectively. The current China market is similar (Section 4.1): four largest CDNs dominate the market.

Starting from 2001, some ICP sites contracted with multiple CDNs simultaneously [9]. Multi-CDN has become a trend recently [14]. The average number of CDNs per top site, in 2013 China, is around 1.7 (Section 4.1).

Another trend is that gigantic ICPs (e.g. Google) tend to build their own CDNs. Once upon the largest CDN customer in North America, now over 70% of Netflix traffic comes directly from Netflix-distributed caches or other infrastructure with Netflix datacenters [17]. The same is happening in China: all major video content providers in China have turned to their own infrastructures (Section 4.2). Large e-commerce ICPs such as Taobao even provides CDN services to their sister companies (Section 4.3).

Streaming [20], given its dominating role in the current Internet traffic, needs specific attention. Netflix used to equally partition its traffic among Akamai, Level 3 and Limelight [17]. While for Hulu, Level 3 gets the biggest share (47%) followed by Akamai (28%) and Limelight (25%) [16]. In China, major video content providers only use CCDN for flash, software distribution and advertisement (Section 4.2).

2.3. Charting

Charting CDN networks is informative to both ICP and CDN providers. Huang *et al.* [10] completely charted Akamai and Limelight networks in 2008. There were around 27 000/6000 content/DNS servers in Akamai, 60% in United States and 90% in top 10 countries. For Limelight, the numbers were 4100/3900 content/DNS servers; 10 data centers resided in USA and nine in other continents. Geographic and ISP distribution of networks were also presented. Later, there are some similar measurement papers [12, 13]: a distinct contribution is that they cluster servers to a set of locations. We perform the same work for the four leading CDNs in China (Section 5.1).

Server diversity has been proved directly related with the network distance between a client and the nearest CDN servers [2]. In addition, Akamai controls the redirection time (around 100 s) to capture the network changes [2].

Based on the charting and performance measurement results, existing and new CDN providers can hypothetically evaluate the benefit and cost of future deployment [10, 12, 13]. Huang *et al.* [10] claimed that Limelight could, by adding just five more data centers, improve its performance to within 10% of Akamai. Triukose *et al.* [13] suggested that Akamai could, without significantly sacrificing the performance, consolidate its hundreds of locations to only 60 data centers. Wang *et al.* [12] stated that for reasonable performance, only 11 locations are required in North America; the number is around 36–72 for global market. The CDN optimization is beyond the scope of this paper.

2.4. Performance

The performance of CDNs is of great interest to network research community. In 2000, Johnson *et al.* [8] presented the first delay performance measurement work of Akamai and Digital Island. A discovery was that CDN does not choose the best server, but instead avoid choosing the worst one. Krishnamurthy *et al.* [9] conducted a detailed measurement of seven CDNs circa 2001. Huang *et al.* [10] conducted a comprehensive measurement for Akamai and Limelight in 2008. For DNS resolve delay, Limelight was 23% higher than Akamai; while for content delay, Limelight was 114% higher. Our work provides performance information for CDN providers in China (Section 6.1).

There are also measurements on CDN throughput. Triukose *et al.* [13] showed that in around 96% cases, Akamai was better in download compared with ICP's original servers; in around 41% cases, Akamai was even at least five times faster. Wang *et al.* [12] used a JavaScript-based tool for client-side download throughput measurement for two major CDNs. They found that two CDNs are

DOI: 10.1002/cpe

Figure 1. Active measurement infrastructure.

comparable in North America, but one is significantly better in Europe and Asia. Liu *et al.* [14] used PlanetLab nodes to measure the download performance of several pay-as-you-go CDNs (e.g. CloudFront) in different time and locations. In this paper, we conduct similar measurements with a focus on comparing CDNs in different regions and ISPs of mainland China. Specific to streaming, Adhikari *et al.* [15] found that all three CDNs have similar overall performance. The daily and instantaneous bandwidth variations were also measured.

2.5. Content

Content types served by CDN is mostly ignored by the community. In 2001, 96–98% served CDN objects were images, which also accounted for 40–60% bytes [9]. At that time, less than 1% objects were streaming, accounted for around 20% bytes. Furthermore, the content in 2001 is almost static: only 2% changed over a 60-day period [9]. An interesting measurement demonstrates the content type of a CDN for content sharing [21]. In this paper, we present a detailed analysis of the content types in China (Section 4.2), which exhibits significant differences.

3. MEASUREMENT OVERVIEW

3.1. Measurement infrastructure

We conduct active measurements to evaluate the performance of CCDN providers in China and their share of traffic volume. Figure 1 shows the measurement nodes deployed throughout the Internet in mainland China. Over 60 measurement nodes are placed in 40 cities spreading to 32 provinces and regions for a couple of reasons. First, China's ISPs are structured at the level of provincial districts. Second, the Internet user population is clustered around a number of large metropolitan areas such as Beijing, Shanghai and Shenzhen. Our measurement nodes cover the full range of ISPs and large metropolitan areas, providing the most representative network access scenarios.

3.2. Measurement methodology

The objectives of the active measurements are to (1) quantify the usage of CDNs by the top ICPs in China and (2) measure the performance of the CCDN providers. Toward these goals, our measurements and analysis follow the procedure illustrated in Figure 2.

1. *Crawler script design and deployment:* we design a crawler script that extracts web links from a given top level website. The script is deployed on all the active measurement nodes and programmed to extract web links from China's top 100 websites. Starting from the top level web page of a site, up to three levels of web links are followed and recorded to reduce the load of the measurement nodes.

DOI: 10.1002/cpe

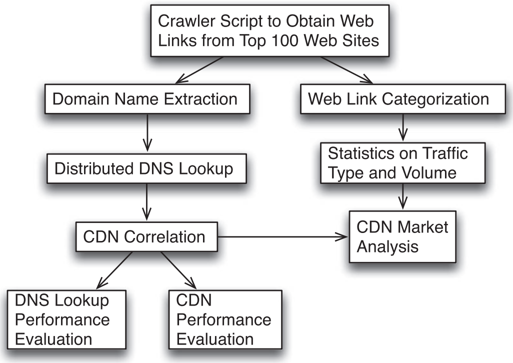

Figure 2. Measurement and analysis workflow.

Table I. Categories of top 100 Internet content providers under study.

Year	News	Video	E-Com	Game	SNS	Search	Other
2013	22	12	28	6	13	0	19
2011	15	10	19	6	10	12	28

SNS, simple notification service; E-Com, e-commerce.

2. *Domain name extraction:* we extract domain names from the collected web links, eliminate redundant ones and prepare unique domain names for DNS lookup.
3. *Web link categorization:* we remove the redundant links from the ones extracted by the spider scripts and group them by ICPs. These categorized links are later used in CDN market analysis.
4. *Statistics on traffic type and volume:* we obtain statistics on the web objects by their types (file, image, video, etc.) based on the suffix of the links. Traffic volume is also recorded for each link and later used in market share analysis.
5. *Distributed DNS lookup and analysis:* we utilize the nearest public DNS server of each measurement node to resolve all the domain names to obtain their CNAME, IP addresses and NS information, which are then sent to a centralized data analysis node to remove redundancy.
6. *CDN correlation:* the DNS analysis results are then compared against a CDN domain name table, to correlate the domain names of original web links with CDN providers that serve the contents. Specifically, we aim to understand, for example what IP addresses are used by which CDN providers and what object types are served by which CDNs for which ICPs.
7. *CDN market analysis:* we study the status and trend of CDN market in China based on the web link categorization and CDN correlation results.
8. *CDN performance analysis:* using the IP addresses resolved from distributed DNS lookup, we use performance measurement tools to evaluate packet delay, loss rate and DNS lookup time.

3.3. Summary statistics

The ICPs focused in this paper are the top 100 websites ranked by Alexa [19] as of June 2011 and June 2013. As listed in Table I, the top 100 websites are classified to the following categories: news, video-serving, e-commerce, online gaming, social and others. We exclude certain types of websites such as search engines and banks from our June 2013 study. This is because (1) we find that these websites mostly use their PCDNs where dynamic data from query transactions are the primary contents and (2) our focus in this paper is to better understand *commercial* CDNs' trend and behavior. While the composition of the websites in each category has changed from 2011 to 2013, these popular websites under study enable us to investigate how representative ICPs leverage CDN capabilities in China.

Based on the methodology described in Section 3.2, our large-scale measurement experiments carried out in both 2011 and 2013 provide a multidimensional view of the state and trend of commercial and customized CDN services used by China's ICPs (Table II). The data analysis on the

DOI: 10.1002/cpe

Table II. Scale of dataset.

Year	Web link	Domain names	Wildcard removal	Distinct IPs	Distinct domain	IPs owned by CCDN
2013	7987632	124355	118327	20110	4125	6909
2011	1225909	2965891	392051	16066	6588	5459

IP, Internet provider; CCDN, commercial content delivery network.

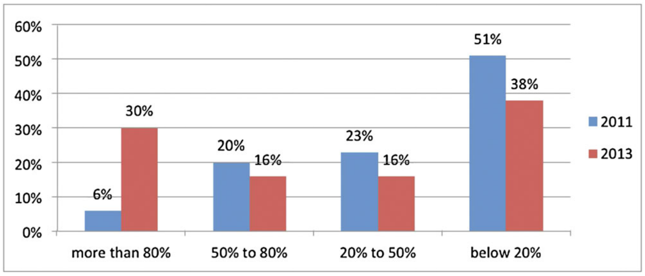

Figure 3. Traffic coverage percentage of commercial content delivery networks in top 100 Internet content providers.

	CDN	number of ICPs served		number of domains served	
		2011	2013	2011	2013
1	ChinaCache	40	45	690	735
2	WangSu	30	41	355	411
3	DiLian	18	25	155	184
4	KuaiWang	8	11	182	129
5	ShiJIHuLian	3	0	8	0
6	TaoBao	5	6	24	53
7	QQ	3	3	43	20

Figure 4. Number of Internet content providers served by content delivery networks.

temporal dimension, that is metrics comparison between 2011 and 2013, is particularly interesting in revealing the trend of China's CDN industry.

4. UNDERSTANDING THE CDN MARKET CHOICE IN CHINA

4.1. CCDN market share

It is observed in Figure 3 that the number of ICPs using CCDNs increases from 59 of the top 100 ICPs in 2011 to 75 ICPs in 2013, as a result the network traffic volume carried by CCDNs increases too. In 2011, 26% of the top 100 ICPs that use CCDN services have more than half of their traffic carried by CCDNs. In 2013, the percentage changes to 46%.

During the 2-year period of time, the major providers and market share of CCDNs have changed. The market share comparison data are shown in Figure 4. Ranked by ICPs served, the top CCDN providers in 2011 are all CCDN, that is ChinaCache, WangSu, DiLian, KuaiWang and Shijihulian. However, in 2013, the top five CDN providers for top 100 ICPs change to ChinaCache, WangSu, DiLian, KuaiWang and TaoBao. It is worth noting that the new player TaoBao is traditionally an e-commerce platform and has now built a CDN to serve other ICPs. Data show that the largest simple notification service portal of China, that is QQ, also provides CDN service for other ICPs. In 2013, QQ provides CDN service for three ICPs in top 100 ICPs.

Such new expansion trend has significant implications to both existing and new CDNs providers. Some existing CCDN providers are losing ground to emerging providers. Meanwhile, the market share is increasingly concentrated to the largest provider ChinaCache.

We use delivery bytes to calculate the market share among those major CCDN providers. Figure 5 illustrates the market share in 2011 and 2013. The trend is that the market is increasingly dominated by a few large scale CDN providers such as ChinaCache and WangSu: ChinaCache increases its

DOI: 10.1002/cpe

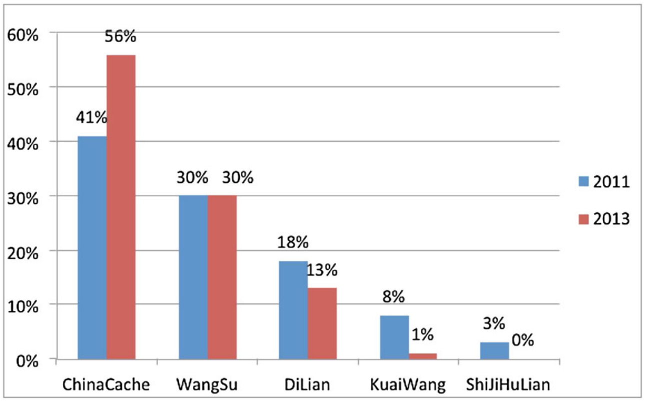

Figure 5. Market share percentage among major commercial content delivery networks.

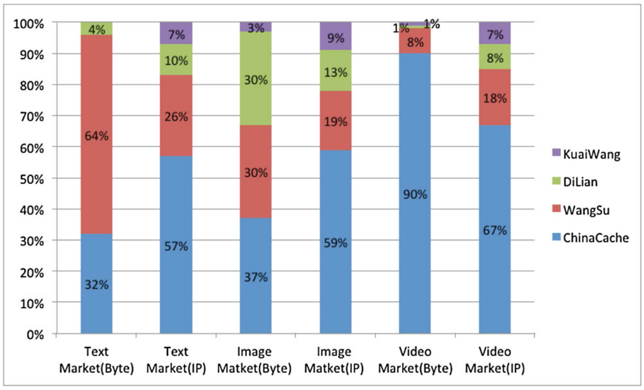

Figure 6. Share percentage in term of market segmentation.

share from 41% to 56% and WangSu preserves its market share of 30%. Smaller CCDN providers' share drops.

From another aspect, we also can find a similar trend. The ICPs tend to choose fewer CCDN providers. In 2011, an ICP choose 2.7 CCDN providers on average, whereas that number drops to 1.7 in 2013. This vendor affinity trend implies that the CCDN market matures into a stable state, where four major CCDNs almost occupied the whole CCDN market.

4.2. *The content types served by CDNs*

In order to investigate the competitiveness of four major CCDNs in 2013, we study the market share percentage among four major CCDNs in 2013 in terms of serving different content types, which is illustrated in Figure 6. It shows that WangSu takes a predominant position in text delivery market, while ChinaCache tops in video market. In image delivery market, ChinaCache, WangSu and DiLian almost fair share the market.

The majority of the data served by CCDNs are via hypertext transfer protocol (HTTP). Six categories of ICPs are among Alexa's top 100 websites: news, video, e-commerce, game, social network and others. After collecting the links by our crawler scripts, we measure the type and size of the content behind every link. Seventy-five ICPs use the service of CCDNs. All ICPs acquire services for textual data, 100% use services for image delivery and nearly 76% for video content delivery.

We conduct in-depth study on the video delivery services by CCDNs. There are 12 video portals, who provide professional video service, among the top 100 ICPs in 2013. All of them build PCDN to serve themselves, and only half of them buy very small portion of video service from CCDNs, which mainly deliver the following four types of contents: 1) small flash animation, 2) audio player, 3) video player and 4) video advertisement. Further analysis on 200 video link samples show their

DOI: 10.1002/cpe

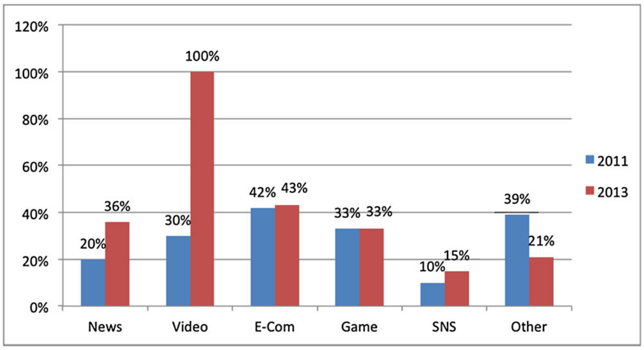

Figure 7. Private content delivery networks by website categories.

composition: 13% flash-based animated advertisement, 47% news and video advertisement, 26% video player and 14% audio player. End users download the audio/video players from CCDNs, which then redirect the user requests to ICP's PCDNs to obtain the content. Most of other non-video ICPs host advertisement/news videos and flash files at CCDNs and use their HTTP service to deliver such contents.

Transactions carried out for e-commerce websites are primarily supported by themselves or Taobao, the largest e-commerce platform in China. Most e-commerce sites only rely on CDNs for limited web or video services. Overall, the content type served by CCDNs is concentrated on HTTP related static data.

We further look into the efficiency of CCDN providers by calculating how they utilize IP addresses to serve different types of contents. The correlations of their market share and infrastructure scale are analyzed in Section 6.2.

4.3. ICPs' PCDNs

It can be seen in Figure 7 that it is a growing trend that ICPs start building their own CDN infrastructures and use CCDNs as supplements. In 2011, 41 ICPs have built customized CDNs and 10 of them are search portals, which are excluded from our measuring target in 2013. In 2013, 44 ICPs built PCDN.

Among the six categories, compared with 2011 building-customized CDNs gets more popular among news, video, e-commerce and social network websites, whereas gaming and other websites do not see an increase. The top gaming sites such as 4399.com primarily offer games for download and play, and other websites such as soufun.com and pcauto.com.cn are mainly information services. Because of the nature of their contents (static text and images) and scale, these two categories of websites simply use CCDN rather than building PCDNs.

Major ICPs build customized CDNs not only for themselves, but also for other ICPs, which have close business relationships with them. Table. III summarizes such relationship observed from measurement results. The customers of Taobao's CDN, including *etao* (Taobao's e-commerce platform), *tmall* (sister company), *Alibaba* (principal company), *cnzz* (sister company), *xiami* (subsidiary) and *alipay* (payment platform), are all subsidiaries of Alibaba or have direct relationship with Taobao. The ICPs using QQ's CDN, including *pengyou* and *paipai* both owned by QQ. Similar use cases are observed for *Sina* with its subsidiary *Weibo*, *Netease* with its subsidiary *youdao.com* and *Sohu* with its *focus.cn*. Therefore, such CDN providers are viewed as ICPs with PCDNs.

Existing ICP such as Taobao and QQ have gained technical and knowledge advantages over the recent CCDN providers. As a result, static HTTP contents are served primarily by commercial providers, whereas the revenue generating services such as e-commerce transactions and videos are handled by ICP themselves.

4.3.1. Case study: QQ.com. QQ.com is the main web portal of the largest online messaging system in China. In 2011, QQ.com has produced 7589 IP addresses, among which only 626 or 6.9%

DOI: 10.1002/cpe

Table III. Customers of private content delivery networks built by ICPs.

ICPs serving	ICPs served
taobao.com	etao.com, tmall.com, alibaba.com, xiami.com, alipay.com, cnzz.com
qq.com	51buy.com, paipai.com, pengyou.com
sohu.com	focus.com, 17173.com
xunlei.com	duowan.com, tgbus.com
sina.com.cn	weibo.com
163.com	youdao.com

ICP, Internet content provider.

Table IV. Content servers deployment.

CDN	# of IPs	# of Clusters	Average # of IPs	China UNICOM	China TELECOM	China MOBILE
ChinaCache	2788	86	32.42	58%	41%	1%
WangSu	872	118	7.39	64%	35%	1%
KuaiWang	260	37	7.02	60%	40%	0%
DiLian	1180	66	17.88	59%	39%	2%

CDN, content delivery network; IP, Internet provider.

are from CCDNs. It has 1180 subdomain names and only 65 utilize CCDNs. In 2013, among the 3406 distinct IP addresses resolved from web links, CCDNs provide 194, that is 5.8%. Only six subdomain names among the total of 1523 are served by CCDN providers.

5. CHARTING THE CDNS

In this section, we apply the charting methodology to four leading CDNs in China, namely, ChinaCache, WangSu, KuaiWang and DiLian. Both content server networks and their DNS infrastructures are charted.

5.1. Content servers

Shown in Table IV are the results of charting and clustering. It is apparent that all providers adopt the Akamai-like *enters deep into ISPs* approach: the average number of IPs per cluster is at most tens. Among three major China ISPs (i.e. UNICOM, TELECOM and MOBILE), the server distributions of these CCDNs are similar: around 60% in UNICOM, 40% in TELECOM and a negligible number of servers in MOBILE. As we will discuss in Section 6.3, such phenomena are largely attributed to the network performance of ISPs: China's UNICOM does not offer as good performance as other ISPs, thus CCDNs utilize more resources to compensate the network performance disadvantages.

Regarding the number of clusters, WangSu is geographically more distributed than others, although ChinaCache has the largest number of IPs. To study the specifics of geographical distribution, we locate the ChinaCache and WangSu clusters precisely. Shown in Figure 8, the blue nodes belong to ChinaCache, while the red nodes belong to WangSu. WangSu reaches every province, including Tibet and Xinjiang, which have very sparse population.

Given the number of IPs and clusters, it should be noted that the distribution of IPs versus clusters is uneven. Shown in Figure 9 is the number of IPs of each CDN in China's provinces. One of the major driving forces for such skewed distribution is the headquarter location: ChinaCache in Beijing, DiLian in Shanghai and Taobao in Zhejiang. ChinaCache tends to build large clusters outside its headquarter: Guangdong, Jiangsu, Shandong, Tianjin and Shanghai. Also headquartered in Shanghai similar to DiLian, WangSu intends to spread its servers to its more geographically distributed clusters.

Compared with 2011, the IPs used by CCDNs in 2013 increase from 5459 to 6909. This trend is not uniform in every province. As illustrated in Figure 10, some provinces see an increase

DOI: 10.1002/cpe

Figure 8. Deployment of ChinaCache and WangSu.

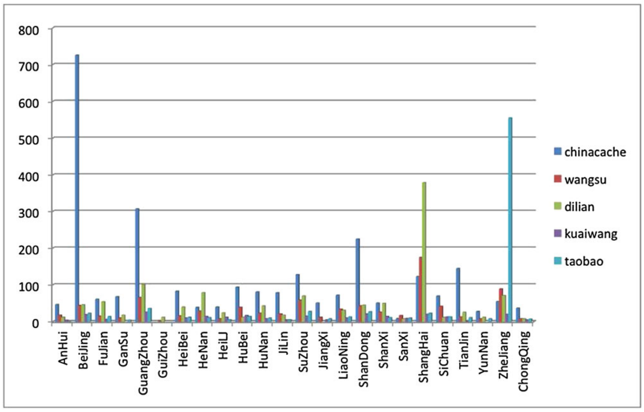

Figure 9. Deployed Internet providers of each content delivery network by provinces.

(e.g. Shanghai and Zhejiang), while others see a decrement (e.g. Beijing and Guangdong). As mentioned in Section 4, the coverage hence the traffic of CCDNs increases substantially.

5.2. DNS infrastructure

All charted CCDN providers use two-tier infrastructure: the top level domain servers respond with a list of recommended second level domain servers; and one of the second level name server eventually answers the query of a CNAME.

Among the CCDNs, ChinaCache has a unique feature: separation of domain name servers. ChinaCache has separate top level domain servers for major ISPs (i.e. China Telecom and China Unicom). Most large ICPs choose to divide users to their ISP-specific ChinaCache top level DNS, while most small ICPs do not. We conjecture the reason being large ICPs use this approach to reduce the round-trip time between users and the top level name servers. For the returned lists of the second level name servers, ChinaCache only returns name servers from the same ISP. Besides, they are sorted based on geographical distance. For example, a local domain name server from Shandong would get a second level name server in Shandong as the first item in the list. We assume the reason is that most query implementation would take the first server in the returned list as the default choice; ranking them in sorted list would reduce DNS query latency.

In contrast, WangSu and other CCDNs have only one set of top level domain servers: users from different ISPs are directed to the same set. What's more, for the returned lists of the second level

DOI: 10.1002/cpe

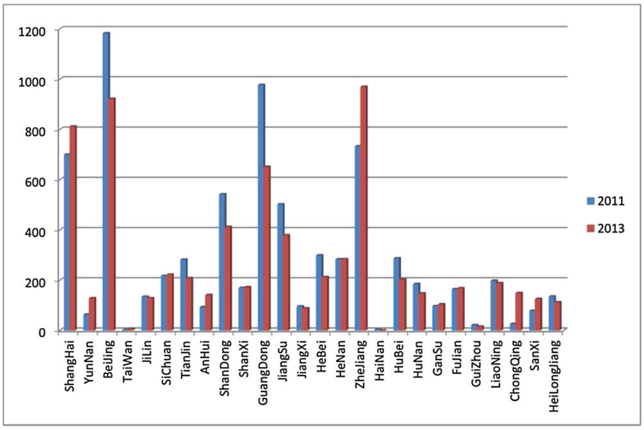

Figure 10. Commercial content delivery network Internet providers of year 2011 and 2013 by provinces.

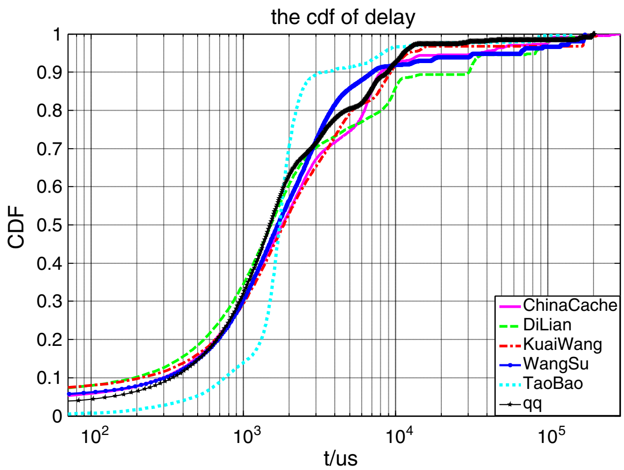

Figure 11. CDF of content delivery network delay.

name servers, WangSu even returns name servers in different ISPs: they are sorted first in ISP order, then in geographical order. The reason to include name serves in other ISPs might be fault tolerance, but we have not confirmed this assumption yet.

6. UNDERSTANDING THE NETWORK AND TECHNOLOGIES OF CDNS IN CHINA

6.1. QoS differences among CCDNs

We evaluate the QoS of CDN services (commercial and self-built) using three metrics: delay, packet loss ratio and DNS lookup time. The following CDN services are selected in the measurement study: ChinaCache, WangSu, DiLian and KuaiWang as CCDN providers and QQ and Taobao as PCDNs built by ICPs. To ensure the accuracy of the measurement, we group the IP addresses of CDNs based on provincial districts and ISPs, then use the nearest measurement node to issue *ping* commands to measure delay and losses. Statistical mean values from ping's output are recorded and used for further analysis.

It shows that CDNs built by ICPs outperform CCDNs in terms of delays. Figure 11 illustrates the CDF of the CDN delays. The services delivering less than 10 ms delay are Taobao (95%), QQ (92%), ChinaCache (92%), WangSu (91%), KuaiWang (88%) and DiLian (82%), ranked by probability. Between the two ICPs (QQ and Taobao), Taobao, an e-commerce platform, has better

DOI: 10.1002/cpe

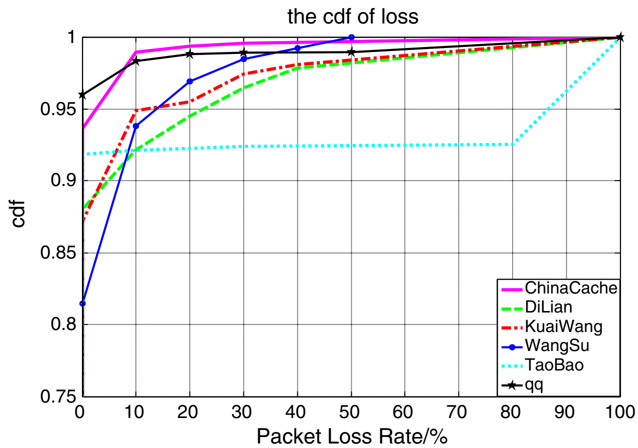

Figure 12. CDF of content delivery network delay packet loss.

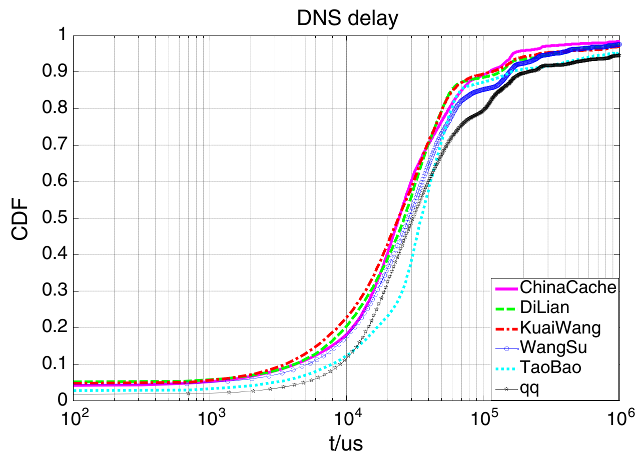

Figure 13. DNS delay CDF.

performance largely because of the real-time requirements of transaction oriented web access by e-commerce customers.

Figure 12 shows that ChinaCache, the largest CCDN, performs well on the loss rate as its zero-loss probability is over 97% because of its large clusters on an ISP network with better performance. The PCDN by QQ has the highest probability of loss rate lower than 10%, outperforming other CCDNs. The PCDN of another ICP under study, Taobao, experiences the lowest probability of a loss rate less than 10% and 7% of probability of a packet loss rate larger than 80%. The primary reason is that Taobao deploys the majority of its CDN servers (555 out of 749 IPs) in Zhejiang Province, where the e-commerce transaction services are carried out, and zero packet loss rate is observed, whereas its servers at other regions of the country, carrying other services such as download, incur high packet losses. Using 2% as the threshold of packet losses suggested by China's telecommunication regulations, the ranking of the CCDNs are QQ, ChinaCache, Taobao, DiLian and WangSu.

DNS delays are measured with a similar approach as in [10] by performing DNS lookups on CDNs' domain names from the measurement nodes deployed nationwide, and the results are depicted in Figure 13. It can be observed that the lookup time concentrates between 40 and 100 ms. The lookup time less than 100 ms are ranked as follows: ChinaCache (88.9%), KuaiWang (88.7%), DiLian (88.3%), Taobao (87.1%), WangSu (85%) and QQ (80%), ordered by probability. The CCDNs outperform ICP's PCDNs because their technology advantages on DNS-based request routing and their understanding of the network architecture.

DOI: 10.1002/cpe

Table V. Infrastructure efficacy (higher is better).

	ChinaCache	WangSu	DiLian	KuaiWang
Text	0.56	2.46	0.40	0
Image	0.63	1.58	2.30	0.33
Video	1.34	0.44	0.13	0.14

Table VI. Number of leading provinces for commercial content delivery network latency.

	ChinaCache	WangSu	DiLian	KuaiWang
TELECOM	6	5	5	0
UNICOM	4	2	1	3

6.2. Correlation between market share and infrastructure scale

We analyze the infrastructure efficiency of CCDN providers by correlating their IP addresses with content types they serve. This is because the IP addresses allocated by ISPs incur additional costs, and the number of IP addresses used by a CCDN implies the scale of its infrastructure. Figure 6 shows that WangSu occupies 64% of the share on text content but uses only 26% of the IP addresses. In contrast, ChinaCache consumes 57% of the IP resources for 32% of the market share. Such resource utilization efficiency suggest the technical competitiveness, thus we define a metric E as $E =$ *(market share) / (IP addresses share)* to evaluate the infrastructure effectiveness of CCDN providers. Larger E means a higher market share over unit resource. As shown in Table V, WangSu has the highest efficiency in text category, DiLian in image category and ChinaCache in video category.

6.3. Correlation between QoS and ISPs

In order to better understand the correlation between CCDNs and ISPs, we conduct QoS measurement (i.e. delay) of different ISPs. We firstly consider the whole China, as treated by most of Chinese CCDNs, into eight major districts, that is Beijing, Shanghai, Guangzhou, Wuhan, Nanjing, Shenyang, Chengdu and Xian. Then, we use our measurement servers to *ping* 23 million IP addresses collected through the work in [22], which represent the distribution of Chinese Internet users. We find that TELECOM has a predominant QoS advantage in six (Shanghai, Guangzhou, Wuhan, Nanjing,Chengdu and Xian, all located in south of China) of eight districts and has larger network coverage for Internet users than UNICOM and other ISPs. Referring to Table IV in Section 5, four largest CCDN providers all deployed more servers in China UNICOM and try to utilize more overlay resources to compensate the network performance disadvantages in UNICOM.

We measure the network latency between a measurement node and a CDN server, both of which reside in the same province and the same ISP, in order to emulate a typical service scenario. Such a measurement is performed for every CCDN provider and in all 26 provinces, and the number of provinces where each CCDN tops others is listed in Table VI. The observation is consistent with the charting results in Section 5.2. Headquartered in south China, WangSu and DiLian have advantage mainly in TELECOM. Headquartered in north China, KuaiWang focuses mainly in UNICOM. ChinaCache's performance is more balanced: six in TELECOM and four in UNICOM. ChinaCache also has the largest number of leading provinces: 10 out 26, or 38% among all provinces.

7. SUMMARY AND IMPLICATIONS

The CDN trend in China is clear. Large ICPs tend to let CCDNs serve static contents such as text and image. For high value services (e.g. streaming) and high QoS demanding services (e.g. transaction services in e-commerce), ICPs tend to build and use their PCDNs and let CCDNs handle the affiliate contents such as media player software and advertisements. The long-term threats faced by a CCDN provider come from not only its competitors but also the large ICPs.

DOI: 10.1002/cpe

Major CCDN providers share much similarity in terms of technology and infrastructure. The reason is that they are dedicated (no ISP-related CDNs), and they all serve similar contents. They have the same ‘enter-deep-into-ISP’ infrastructure, and their request routing mechanisms are all DNS-based. The major factor affecting the resource deployment of CCDNs is the underlying ISPs. China UNICOM has less clients and smaller coverage compared with TELECOM, and the network performance of UNICOM is relatively worse. As a result, all CCDNs deploy more resources inside UNICOM to compensate service quality.

Although major CCDNs in China mainly serve static contents, they have distinctions among each other. Their resource utilization efficiencies are different in text, image and video contents. ICPs tend to use different CDNs to deliver different types of contents. As the market share is increasingly centralized to a few largest CCDNs, the average number of CCDNs chosen by an ICP decreases from 2.7 in 2011 to 1.7 in 2013.

The short-term perspective for China’s CCDNs remains optimistic. The service price has dropped due to the competition among CCDNs and the development of technologies. As a consequence, a growing number of ICPs purchase services for static contents. This trend is evident through the increase of CCDN usage by top ICPs from 2011 to 2013. Because ICPs gradually use their PCDNs for high value/volume and high QoS demanding services such as video, the long-term development of CCDNs should focus on innovations from both technologies and business models. Emerging CDN brokers may become one of the directions [7, 14].

ACKNOWLEDGEMENTS

The authors would like to thank anonymous reviewers for their valuable comments. This work is supported by the ‘National Basic Research Program of China (973 Program) (no. 2011CB302601 and no. 2012CB315801)’, by the ‘National Natural Science Foundation of China (no. 61222213, no. 61170290, and no. 61202107)’, by the ‘EU FP7 CLIMBER project under Grant Agreement (no. PIRSES-GA-2012-318939)’, and by ‘Jiangsu International Cooperation Program of Science and Technology (grant no. BZ2013018)’.

REFERENCES

1. Labovitz C. *CDN and Over-The-Top Traffic Data, Content Delivery Summit* , 2012. (Available from: http://www.contentdeliverysummit.com/2012/) [Accessed on December 30, 2014].
2. jan Su A, Choffnes DR, Kuzmanovic A, Bustamante FE. Drafting behind akamai (travelocity-based detouring). *Proceedings of ACM SIGCOMM 2006*, 2006; 435–446.
3. Cohen J, Repantis T, McDermott S, Smith S, Wein J. Keeping track of 70,000+ servers: the akamai query system. *Proceedings of the 24th Usenix Large Installation System Administration Conference (LISA)*, 2010.
4. Nygren E, Sitaraman RK, Sun J. The akamai network: a platform for high-performance Internet applications. *ACM SIGOPS Operating Systems Review* 2010; **44**(3).
5. Repantis T, Cohen J, Smith S, Wein J. Scaling a monitoring infrastructure for the akamai network. *ACM SIGOPS Operating Systems Review* 2010; **44**(3).
6. Wang Y, Tian C, Jiang H, Liu X, Chen J, Liu W. Sharp: a scalable framework for dynamic joint replica placement and request routing scheduling. *Global Telecommunications Conference (GLOBECOM 2011), 2011 IEEE*: IEEE, 2011; 1–5.
7. Liu X, Dobrian F, Milner H, Jiang J, Sekar V, Stoica I, Zhang H. A case for a coordinated Internet video control plane. *Proceedings of the ACM SIGCOMM*, 2012.
8. Johnson KL, Carr JF, Day MS, Kaashoek MF. The measured performance of content distribution networks. *Computer Communications* 2000; **24**(2).
9. Krishnamurthy B, Wills C, Zhang Y. On the use and performance of content distribution networks. *Proceedings of the ACM SIGCOMM Internet Measurement Workshop 2001*, 2001.
10. Huang C, Wang A, Li J, Ross KW. Measuring and evaluating large-scale CDNs. *Proceedings of the 2008 ACM SIGCOMM Conference on Internet Measurement Conference*, 2008.
11. Ager B, Mühlbauer W, Smaragdakis G, Uhlig S. Web content cartography. *Proceedings of the 2011 ACM SIGCOMM Conference on Internet Measurement Conference*, 2011.
12. Wang YA, Huang C, Li J, Ross KW. Estimating the performance of hypothetical cloud service deployments: a measurement-based approach, 2011.
13. Triukose S, Wen Z, Rabinovich M. Measuring a commercial content delivery network. *Proceedings of WWW 2011*, 2011.
14. Liu HH, Wang Y, Yang YR, Wang H, Tian C. Optimizing cost and performance for content multihoming. *Proceedings of the ACM SIGCOMM 2012*, 2012; 371–382.

DOI: 10.1002/cpe

15. Adhikari VK, Guo Y, Hao F, Varvello M, Hilt V, Steiner M, Zhang ZL. Unreeling netflix: understanding and improving multi-CDN movie delivery. *Proceedings of the IEEE INFOCOM 2012*, 2012.
16. Adhikari VK, Guo Y, Hao F, Hilt V, Zhang ZL. A tale of three CDNs: an active measurement study of Hulu and its CDNs. *Proceedings of the IEEE INFOCOM 2012 GIS Workshop*, 2012.
17. DeepField. *First data on changing netflix and CDN market share*, 2012. (Available from: http://www.deepfield.com/2012/06/first-data-on-changing-netflix-and-cdn-market-share/) [accessed on 30 December 2014].
18. NetFlix. *Announcing the netflix open connect network*, 2012. (Available from: http://blog.netflix.com/2012/06/announcing-netflix-open-connect-network.html) [accessed on June 2014].
19. Alexa. *Alexa's top 100 site ranking for China*, 2013. (Available from: http://www.alexa.com/topsites/countries/CN) [accessed on June 2014].
20. Tian C, Alimi R, Yang YR, Zhang D. Shadowstream: performance evaluation as a capability in production internet live streaming networks. *Proceedings of the ACM SIGCOMM 2012 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communication*: ACM, 2012; 347–358.
21. Yu T, Tian C, Jiang H, Liu W. Measurements and analysis of an unconstrained user generated content system. *2011 IEEE International Conference on Communications (ICC)*: IEEE, 2011; 1–5.
22. Yin H, Chang H, Liu F, Zhan T, Zhang Y, Li B. Discovering a large scale Internet topology: complementary and contrast view. *IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications*, 2012; 1491–1498.

DOI: 10.1002/cpe