Cloud based Dynamically Provisioned Multimedia Delivery: An Elastic Video Endpoint

Alistair Blair, Gerard Parr, Philip Morrow, Bryan Scotney and Aaron McConnell
India-UK Centre of Excellence for Next Generation Networks
Faculty of Computing and Engineering, University of Ulster,
Coleraine, Northern Ireland
blair-a6@email.ulster.ac.uk,{gp.parr, pj.morrow, bw.scotney,a.mcconnell}@ulster.ac.uk

Steve Appleby and Mike Nilsson
Video Delivery Research,
BT Innovate & Design, Adastral Park,
Ipswich, England
{steve.appleby, mike.nilsson}@bt.com

***Abstract*—Content Delivery Networks are commonplace in today's Internet and are an important technique in the distribution of multimedia content to the plethora of Internet Protocol enabled devices. However, it has been recognised that current networks are many times over provisioned server side for peak demand and therefore greatly under utilised at other times. The emergence of cloud computing as a commercial reality has created the opportunity where content delivery networks can leverage the resources of existing cloud providers to increase capacity when required. In this paper, we propose an Elastic Video Endpoint (EVE), a virtualised multimedia distribution resource, which can utilise cloud resources to dynamically provision capacity in real time. Initial results have shown that the system can respond to increased load and provide extra bandwidth capacity on demand.**

Keywords-cloud; elastic; content delivery network; dynamic provisioning.

I. Introduction

It has been predicted that video traffic will account for around 90% of the 966 exabytes of global Internet Protocol traffic that will cross the globe in 2015 [1]. The high bandwidth and strict Quality of Service (QoS) requirements such as lower start up delay, reduced end-to-end delay and higher continuity of multimedia, creates many challenges in the area of content management and content delivery across the Internet. Degradation of any of the above factors can have adverse effects on a user's Quality of Experience (QoE), which in turn can lead them to complain or change the service provider they are using.

In an attempt to combat these challenges, clusters of machines connected to the Internet, containing replica data can be strategically placed at various geographic locations to improve dissemination performance. These Content Delivery Networks (CDN) [2], [3] offer a good way to decrease core network bandwidth, reduce network latency and lower delivery costs.

Typically a CDN will carry out the following functions [4]:

- Performs redirection of connection requests to the nearest suitable surrogate server, when a user attempts to download content;
- Provides the ability to deliver various content from a set of surrogate servers that are placed at various geographic locations;
- Perform content outsourcing to control the content that is stored on the surrogate servers that form the CDN and how it is replicated from the source server;
- Provides management services that monitor and store data on requests, cache hit/misses and accounting of content usage.

There are a number of variants of these commercially available content delivery networks and these can be categorised as:

- Highly distributed, e.g., Akamai [5], rent or place servers in the data centres of many Internet Service Providers (ISPs) around the world;
- Big Data, e.g., Limelight [6], build and run their own data centres around the world;
- P2P Assisted, e.g., Bittorrent [7], share content in a collaborative from many different sources, users, web caches and proxies;
- Cloud, e.g., Amazon CloudFront [8], enables content providers to provision capacity from Amazons cloud resources in a pay as you go manner;

However, it has been noted that content delivery networks, in any guise, are many times over provisioned and therefore under utilised [9]–[15]. This over provisioning means that that the CDN infrastructure is expensive to implement and manage, [16], [17]; however, it is required due to the lack of overload protection, so that flash crowds can be dealt with effectively. To reduce this provisioning, would increase the risk of lowering the end user experience. The work of Sun et al. [18] has shown that "10% of connections are server-limited at least 40% of the time." Cloud CDNs are a new and emerging approach [12], [19]–[22], that use cloud resources, namely cloud storage to reduce the cost associated with implementing content delivery services. However, again, these resources are created at different locations across multiple clouds and can lead to over provisioning due to the lack of overload protection.

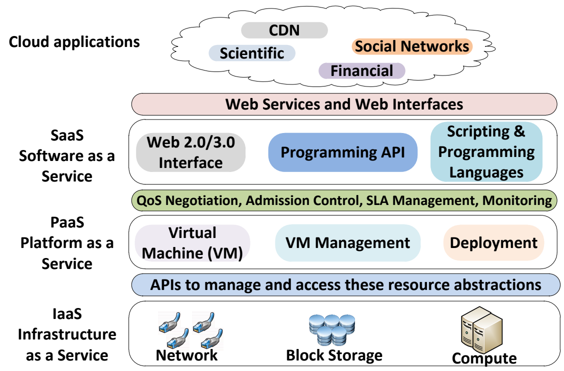

Fig. 1: Cloud Computing layered architecture

The system proposed in this paper uses metrics (CPU, Memory, Disk, Network), obtained from physical machines and hosted Virtual Machines (VMs) in an attempt to dynamically provision resources namely bandwidth when it is required, therefore increasing utilisation while minimising provisioning cost. The paper focuses mainly on Network and CPU metrics. The elastic nature of Cloud resources offer the ability to dynamically provision resources when they are required, this in turn enables resources at a single location to grow when needed to allow higher utilisation across any provisioned hardware.

The remainder of this paper is organised as follows: Section II gives a brief overview of cloud computing and its layered architecture. Section III documents related work in the area of Content Delivery Networks, populars CDNs, integrated cloud based content delivery and details some of the challenges that are being faced. Section IV details the proposed system, and finally, Section V gives some results, discussion and future work.

II. Cloud Computing

The emergence of the cloud computing paradigm as a commercial reality has created a new landscape for Internet based computing, whereby self owned IT resources can be reduced and replaced by the computation-as-a-service model that cloud providers can offer, enabling users to reduce the Capital Expenditure (CAPEX) and Operating Expenditure (OPEX). Virtualisation has allowed cloud providers to offer computing resources (compute, storage and network) as a service that can be dynamically provisioned at multiple geographical locations when required. A cloud platform is typically made up of four distinct layers: data storage, data management, data service, and user access [23].

Fig. 1 gives a representation of the layered structure of the cloud computing architecture, which consists of four layers:

- Infrastructure as a Service (IaaS) or data storage layer provides an abstract view towards the under lying compute, storage and network allowing virtual instances of mini data centres.
- Platform as a Service (PaaS) or data management layer combines infrastructure, operating systems and application software and offers it as a utility.
- Software as a Service (SaaS) or data service layer provides software products and services as a utility that can be used on demand.
- Cloud applications or user access provides the access point of applications to the Cloud.

Due to the high QoS of requirements of applications such as real-time multimedia, cloud computing has become particularly attractive for content delivery. Typically content delivery services are operated using dedicated servers that lack the dynamic nature of cloud computing or cloud storage and fail to fully utilise the elastic abilities that cloud can offer. While the properties described above can be seen as the advantages of cloud computing, there are also some disadvantages, namely resource contention, which occurs when VMs are oversubscribed, i.e., contending for the same physical resources, leading to poor application performance, causing user experience to deteriorate. This paper considers the option of utilising cloud computing resources to operate delivery endpoints that can grow/shrink in real time, when demand for their services changes while maintaining the high levels of QoS that multimedia data requires.

III. Related Work

A. General

Much research has been carried out in the area of data locality and the methods used to disseminate multimedia data to end-users [17], [24]–[27]. There are currently two key distribution techniques used to disseminate media across the Internet, namely CDN architectures and P2P architectures. Recent work has seen attempts to utilise both techniques. TopBT [24] is a topology aware Bittorrent client that can reduce download traffic by 25% while increasing download speeds by around 15%. Alessandria et al. [25] analysed some commercial P2P video applications and found that they were able to cope with impairments caused by delay, packet loss and insufficient bandwidth. However, their study did show that when all peers where affected by bottlenecks they failed to recover. Seyyedi et al. [26], compare connected and unconnected meshes and show that the connected mesh offers significant improvements in end-to-end delay and distortion, while Kang et al. [27] define a hybrid CDN-P2P architecture that allows the CDN network to take the load during quiet periods and when busy the P2P component allows the system to compensate by enabling neighbours to distribute content, relieving stress on the CDN, Tonget al. [17], propose a new web service P2PCDN architecture to help distribute content from under provisioned servers, they believe that their architecture could be provisioned using cloud computing.

The work documented above focuses on using client devices to help in the distribution process, however, these have their weak points, which include high background traffic [28] and energy tradeoffs [9].

B. Popular CDNs

Akamai is currently the market leader of content delivery services, with “nearly one hundred thousand servers, deployed in 72 countries” [5]. These servers are provisioned in many different ISP data centres around the world, whereas Limelight [6] has a few large data centres placed around the globe. However, research has shown that the performance of Akamai could be maintained even if the number of their servers was reduced [29]. Akamai have recently changed their status from being a CDN provider to a cloud provider. The integration of cloud computing is affecting all content providers and at present cloud storage is a significant topic in the research community.

C. Integrated CDN

Cloud computing has created a new concept of Cloud storage, whereby large data stores can be made available to users dynamically when they are required. Cloud storage is very different from traditional storage and whereas traditional storage was of a fixed size, cloud storage has the ability to grow if required. This service is offered on a pay-as-you go basis and so costs can be controlled and managed. In terms of functionality, it is able to deliver a variety of online services, as opposed to traditional storage systems that are aimed at large scale transactional processing and high performance computing. Wu et al. [30], and Huo et al. [31] detail the advantages of cloud storage and the challenges that will face the technology while Lin et al. [19], use cloud storage as the basis of a new content delivery network called CCDN. Simulations by Wang et al. [20], have also shown that cloud storage offers a highly scalable and fault tolerant platform that offers lower delay and higher bandwidth to end users.

D. CDN Challenges

Distribution servers must be over provisioned for peak demand, due to their lack of overload protection (each instance will have limited CPU, Memory, Storage and Network bandwidth). Sun et al. [18], argue that, “passively monitoring the transport-level statistics,” of a server is a much better approach, as the ability to monitor conventional metrics is much too difficult, we would argue that with the advent of virtualisation the ability to monitor core performance metrics offered by the hypervisor, e.g., CPU, Disk, Network and Disk at twenty second intervals is a novel and promising technique (Twenty seconds is the smallest interval offered by VMware). No matter how large the content storage is, bandwidth is required to disseminate the content.

In summary, the above literature makes use of traditional physical resources, while cloud computing has enabled smaller content providers to utilise delivery services that would otherwise be out of their reach, the delivery services are still over provisioned, with multiple instances running at any one location. However, cloud storage is only a minor part of the flexibility that is offered by cloud computing, by utilising cloud computing the ability to dynamically add extra capacity and real-time monitoring of an instance is possible. The ability to monitor a VM and dynamically add/remove extra capacity in real-time presents an opportunity where the foot print of a system can be kept to a minimum, therefore further reducing the financial and data cost associated with cloud based content delivery.

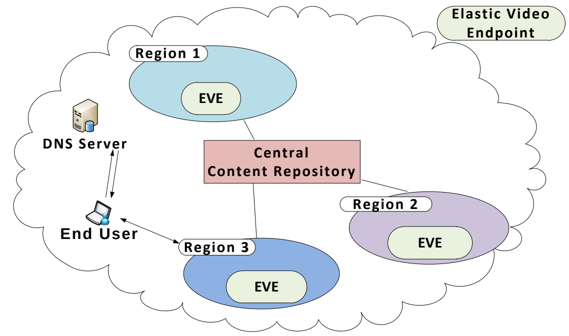

Fig. 2: A dissemination architecture consisting of multiple EVE resources

IV. System Overview

The literature review above has shown that research in the area of content delivery is a popular topic, with developments taking place in all aspects of the concept. However, more recently, the realisation of cloud computing has created a platform that removes the need of content providers to either pay for their own hardware or expensive third party distribution platforms. Content producers can now utilise cloud resources to store and distribute their content in a pay as you go manner. While this has reduced the expense for providers, it still leaves the problem of over provisioning due to the lack of overload protection within these systems. In an attempt to dynamically provision content, while still maintaining a high quality of service we propose a new delivery endpoint, based on cloud resources, “EVE,” Elastic Video Endpoint. In Fig. 2, we can see an architectural overview of a distribution network consisting of multiple endpoints located in different geographical locations. A central content repository retains a copy of all the content that is available for dissemination.

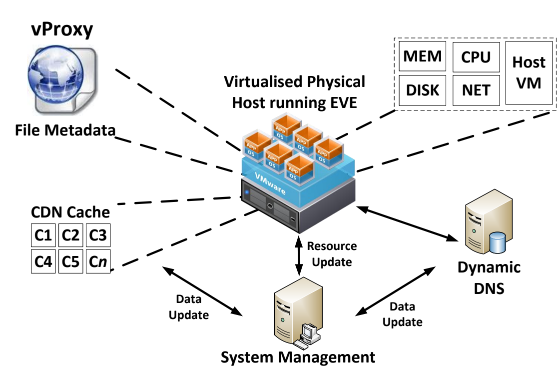

Fig. 3: Components of an Elastic Video Endpoint

From here content can be replicated to endpoints placed in different geographic regions when it is deemed necessary. Each region contains an instance of Elastic Video Endpoint that is adequately provisioned to disseminate content under current conditions. Each endpoint contains a subset of the content that is held at the central repository, which is determined by its current popularity. If the content delivery conditions change i.e., a flash crowd situation occurs where resources become contented, the endpoint can provision extra resources to facilitate the extra demand by utilising the elastic nature of cloud computing. An intelligent endpoint that can scale its resources depending on the load that the system is currently under is shown in Fig 3. The system takes a subset of content from a central master repository and stores it using cloud storage. Current CDNs are optimised for small file delivery [32], [33] and not the large files that are required to store High Definition (HD) and Super HD video. The content used is a mixture of HD and Super HD videos that range in size from a few hundred megabytes to a few gigabytes.

EVE is made up of five major components; these are the VMware Infrastructure, Dynamic DNS, a vProxy, a Cache and System Software.

(a) VMware Infrastructure
VMware infrastructure software has been chosen as the platform on, which to base the cloud platform as it is an industry standard with VMWare controlling eighty percent of the server virtualisation market [34]. The exposed API's of the software provide the ability to access all aspects of the hypervisor and its associated host enabling a custom monitoring system written in C# to be created. This monitoring system can access a large array of metrics. The specific performance metrics to be recorded are CPU, Memory, Disk and Network. These metrics cover many aspects of the system, including the physical host and all VMs running on that host. These metrics give an accurate insight into the current health of both the VM's and the host, using the VMware API's the metrics are available at twenty second intervals. Example metrics are percentage of CPU capacity and throughout of the network.

(b) Dynamic DNS
The request routing component determines the best end-point to facilitate the end users' request. If the content isn't cached locally then the user is redirected to another more suitable location by means of a DNS-based redirect. DNS based requests are highly efficient and help to reduce access time to the endpoint. This enables the system to carry out load balancing so that resource allocation at the endpoint doesn't become overloaded.

(c) vProxy
The proxy keeps a record of files that are stored locally and also of the files that are held remotely. When a cache hit or miss occurs the vProxy updates the relevant record with the information and then redirects the user the local or remote file. By keeping an accurate account of the content that is cached or needs to be cached, the system can fully utilise the space that it has.

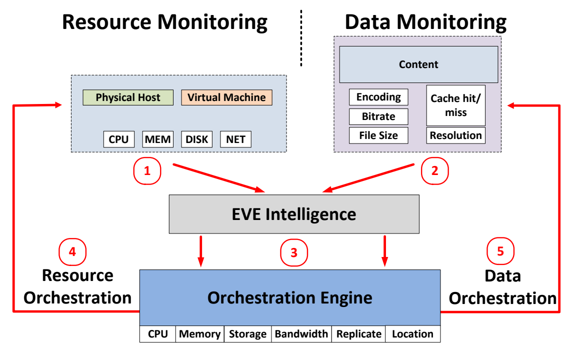

Fig. 4: EVE processes and flow

(d) Cache
The cache is an area of cloud storage that is used to store content for dissemination. The cache is mounted with in the system drive as a folder. This enables the Operating System partition to be minimal in size while the content drive can also be kept to a minimum. This allows the system footprint to remain at a minimum until such times as it needs to expand to hold extra content.

(e) System Software
The management component monitors the remote database to decide when corrective, (e.g., add an extra NIC, increase the cache size or create a new endpoint instance) should be taken to prevent an endpoint from failing. The system takes values from the monitoring database CPU, Memory, Disk and Network, when used along with the cache hits and other content metadata to determine, which file is placing the endpoint under pressure. Corrective action may include increasing capacity or temporarily redirecting future connections to another endpoint via the intelligent DNS.

In order for EVE to operate, there are a number of processes that occur in the general operation and maintenance of each instance. These processes are shown in Fig. 4:

1) Resource monitoring
Performs data acquisition for host and VM metrics for use by the optimiser of metrics such as CPU, Memory, Disk and Network.
2) Data monitoring
Performs data acquisition for use by the optimiser, about content that is hosted at the endpoint, e.g., popularity, cache hits/misses and bitrate.
3) System Intelligence
Performs analysis of the real-time metrics coming from both the data monitoring and resource monitoring processes. Using these metrics the system determines if content should be added/removed or deleted from endpoint instances. Also whether an endpoint instance requires more/less capacity, these decisions are then passed to the relevant orchestration engine for execution.

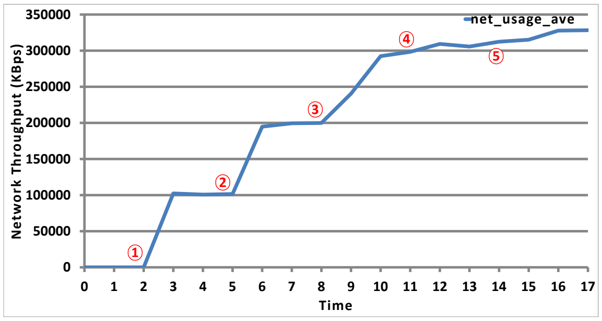

Fig. 5: Network throughput against time

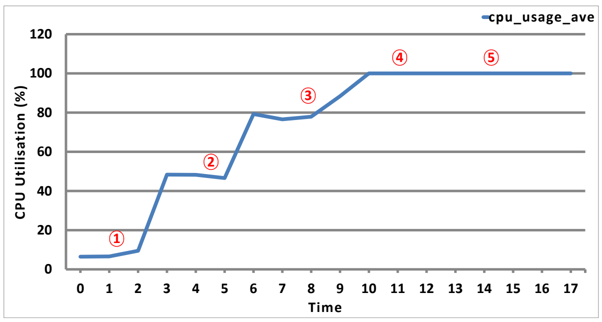

Fig. 6: VM CPU against time

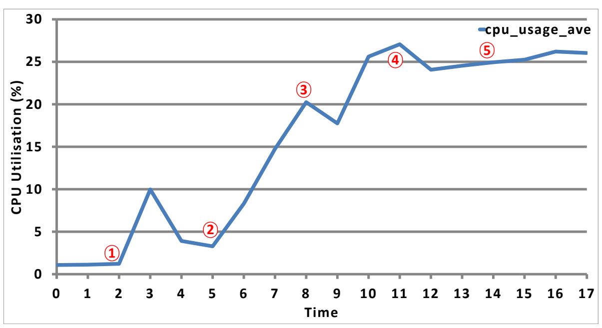

Fig. 7: Host CPU against time

4) Resource Orchestration
Performs dynamic addition or removal of resources resources to a Virtual Machine when and if they are available. If resources are contented on the host then action to create a replica at the same or a different site can be initiated.
5) Data Orchestration
Performs control over the content at the endpoint, to include data replication, deletion or migration, while also updating the dynamic DNS system to provide load balancing across the system.

V. Experiment and Discussion

Experimental Setup

Traditional physically hosted application servers are limited by the resources of that machine, cloud computing enables the dynamic provisioning of extra resources, e.g., bandwidth and storage when they are required. However, even with the ability to add these extra resources a point will arise when the VM is no longer able to meet demand.

The experiment is designed to show that as load on a Virtual Machine increases the resources associated with that VM will also increase until a threshold is reached where adding extra resources will have no or little effect.

In this paper we consider network bandwidth and its effect on CPU. Adding and removing NICs when required to alleviate bandwidth pressure on a VM to disseminate extra data. To demonstrate this a cloud distribution server was created using Windows 2008 R2 web edition with a configuration of 1vCPU, 1024 MB memory, one network interface card and a 40 GB thin provisioned system disk, a second 40GB is mounted for content storage, the entire VM is hosted on a DELL R515 blade server with two Opteron 8 core processors, 16GB of RAM and 12 gigabit ethernet cards, running VMware ESX 5.0.

In order generate a load on the media server, openload [35] was used generate http requests on the IP address assigned to the single NIC. As the load increased it would be expected to see an increase in the CPU utilisation of both the host and the VM. As a link becomes fully utilised, an extra virtual NIC is added and a new instance of openload created against its associated IP address. This process is repeated until a point is reached where the total throughput from the server reaches a maximum, it is expected that the graphs will show that the CPU is the limiting resource.

Discussion

The experiment results showing the network throughput are documented in Fig. 5. Baseline throughput occurs until point 1 where the first instance of openload is initiated. At this point the throughput increases until a maximum is reached, here the throughput levels out. The same point in Fig. 6 shows that the CPU follows a smiler trend. When a second NIC (2) is added and load placed on it, this again results in an increase of throughput with a corresponding increase in the VM CPU. When a third NIC (3) is added the increase achieves a similar addition to the total but takes slightly longer to reach its maximum. This increase in time can be attributed to the CPU reaching 100% utilisation Fig. 6 point (4). Fig. 5 shows extra vNICs being added at point (4) and (5), however, the total throughput levels out with little or no change, due to resource contention on the CPU. At points (2) and (3) we can notice a slight dip in the CPU, we believe this can be attributed to resource discovery as the new vNIC is added to the endpoint.

Fig. 5 and Fig. 6 deal with the Virtual Machine; however, we believe that the host must also be considered, Fig. 7 shows the physical host CPU over the same time period. the graph shows that there is some increase in CPU utilisation on the host as the VM CPU increases. Others spikes in the host CPU could be attributed to other VM's and processes that are running on the host, this information is important has it has an influence on

 ISBN: 978-1-61208-216-5

determining on whether or not resources can be dynamically added to the endpoint at times of VM contention.

VI. Conclusion and Future Work

In this paper, we implemented an initial version of an intelligent multimedia delivery endpoint based on Cloud computing infrastructure called EVE. The results show that the endpoint can provision extra capacity in real time when required, however, it can be seen that resources namely CPU have a limiting factor on the total bandwidth that can be provisioned.

Further work will aim at enhancing the endpoint, to allow the addition or deletion of extra capacity when required. The ability to implement an expanding cache when required and some prediction algorithms to predict provisioning will be developed. This work it is hoped will be detailed in future publications

Acknowledgment

The authors wish to acknowledge the funding received for this project from BT-EPSRC CASE award as part of the India-UK Advanced Technology Centre of Excellence in Next Generation Networks Systems and Services.

References

[1] Cisco visual networking index. [retrieved: April, 2012]. [Online]. Available: http://www.cisco.com/en/US/solutions/collateral/ns341/ns525/ns537/ns705/ns827/white_paper_c11-520862.pdf

[2] Edgecast. [retrieved: April, 2012]. [Online]. Available: http://www.edgecast.com

[3] Cachefly. [retrieved: April, 2012]. [Online]. Available: http://www.cachefly.com/

[4] Hao, "Content delivery networks: a bridge between emerging applications and future IP networks," *Network, IEEE*, vol. 24, no. 4, pp. 52–56, 2010.

[5] Akamai technologies. [retrieved: April, 2012]. [Online]. Available: http://www.akamai.com

[6] Limelight networks. [retrieved: April, 2012]. [Online]. Available: http://www.limelightnetworks.com/

[7] Bittorrent. [retrieved: April, 2012]. [Online]. Available: http://www.bittorrent.com.

[8] Amazon cloudfront. [retrieved: April, 2012]. [Online]. Available: http://aws.amazon.com/cloudfront/

[9] A. Feldmann, A. Gladisch, M. Kind, C. Lange, G. Smaragdakis, and F.-J. Westphal, "Energy trade-offs among content delivery architectures," *Telecommunications Internet and Media Techno Economics (CTTE), 2010 9th Conference on*, pp. 1–6, 2010.

[10] D. Niu, B. Li, and S. Zhao, "Understanding demand volatility in large VoD systems," in *NOSSDAV '11: Proceedings of the 21st international workshop on Network and operating systems support for digital audio and video*. ACM Request Permissions, Jun. 2011.

[11] Z. Liu, M. Lin, A. Wierman, S. H. Low, and L. L. H. Andrew, "Greening geographical load balancing," in *SIGMETRICS '11: Proceedings of the ACM SIGMETRICS joint international conference on Measurement and modeling of computer systems*. ACM Request Permissions, Jun. 2011.

[12] V. Aggarwal, X. Chen, V. Gopalakrishnan, R. Jana, K. Ramakrishnan, and V. Vaishampayan, "Exploiting virtualization for delivering cloud-based IPTV services," in *Computer Communications Workshops (INFOCOM WKSHPS), 2011 IEEE Conference on*, 2011, pp. 637–641.

[13] F. Ramos, R. Gibbens, F. Song, P. Rodriguez, J. Crowcroft, and I. White, "Reducing energy consumption in IPTV networks by selective pre-joining of channels," *Green Networking '10: Proceedings of the first ACM SIGCOMM workshop on Green networking*, Aug. 2010.

[14] N. Xu, J. Yang, M. Needham, D. Boscovic, and F. Vakil, "Toward the Green Video CDN," in *Green Computing and Communications (GreenCom), 2010 IEEE/ACM Int'l Conference on & Int'l Conference on Cyber, Physical and Social Computing (CPSCom*, 2010, pp. 430–435.

[15] I. Vaishnavi, P. Cesar, D. Bulterman, and O. Friedrich, "From IPTV services to shared experiences: Challenges in architecture design," *Multimedia and Expo (ICME), 2010 IEEE International Conference on*, pp. 1511–1516, 2010.

[16] Z. Lu, J. Wu, and W. Fu, "Towards a Novel Web Services Standard-Supported CDN-P2P Loosely-Coupled Hybrid and Management Model," in *Services Computing (SCC), 2010 IEEE International Conference on*, 2010, pp. 297–304.

[17] J. Tong, K. Xu, and R. Pi, "A new Web Service Structure of Combining P2P and CDN Technologies," *Web Society (SWS), 2010 IEEE 2nd Symposium on*, pp. 475–479, 2010.

[18] P. Sun, M. Yu, M. J. Freedman, and J. Rexford, "Identifying performance bottlenecks in CDNs through TCP-level monitoring," in *W-MUST '11: Proceedings of the first ACM SIGCOMM workshop on Measurements up the stack*. ACM Request Permissions, Aug. 2011.

[19] C.-F. Lin, M.-C. Leu, C.-W. Chang, and S.-M. Yuan, "The Study and Methods for Cloud Based CDN," in *Cyber-Enabled Distributed Computing and Knowledge Discovery (CyberC), 2011 International Conference on*, 2011, pp. 469–475.

[20] Y. Wang, X. Wen, Y. Sun, Z. Zhao, and T. Yang, "The Content Delivery Network System Based on Cloud Storage," in *Network Computing and Information Security (NCIS), 2011 International Conference on*, 2011, pp. 98–102.

[21] H. A. Tran, A. Mellouk, and S. Hoceini, "QoE Content Distribution Network for Cloud Architecture," *Network Cloud Computing and Applications (NCCA), 2011 First International Symposium on*, pp. 14–19, 2011.

[22] Y. Wang, C. Huang, J. Li, and K. Ross, "Estimating the performance of hypothetical cloud service deployments: A measurement-based approach," in *INFOCOM, 2011 Proceedings IEEE*, 2011, pp. 2372–2380.

[23] R. Xue, Z.-S. Wu, and A.-N. Bai, "Application of Cloud Storage in Traffic Video Detection," in *Computational Intelligence and Security (CIS), 2011 Seventh International Conference on*, 2011, pp. 1294–1297.

[24] S. Ren, E. Tan, T. Luo, S. Chen, L. Guo, and X. Zhang, "TopBT: A Topology-Aware and Infrastructure-Independent BitTorrent Client," *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9, 2010.

[25] E. Alessandria, M. Gallo, E. Leonardi, M. Mellia, and M. Meo, "P2P-TV Systems under Adverse Network Conditions: A Measurement Study," *INFOCOM 2009, IEEE*, pp. 100–108, 2009.

[26] S. Seyyedi and B. Akbari, "Hybrid CDN-P2P architectures for live video streaming: Comparative study of connected and unconnected meshes," *Computer Networks and Distributed Systems (CNDS), 2011 International Symposium on*, pp. 175–180, 2011.

[27] S. Kang and H. Yin, "A Hybrid CDN-P2P System for Video-on-Demand," *Future Networks, 2010. ICFN '10. Second International Conference on*, pp. 309–313, 2010.

[28] P. Shi, H. Wang, Y. Gang, and X. Yuan, "ACON: Adaptive construction of the overlay network in CDN-P2P VoD system," in *Communication Software and Networks (ICCSN), 2011 IEEE 3rd International Conference on*, 2011, pp. 182–187.

[29] S. Triukose, Z. Wen, and M. Rabinovich, "Content delivery networks: how big is big enough?" *SIGMETRICS Performance Evaluation Review*, vol. 37, no. 2, Oct. 2009.

[30] J. Wu, L. Ping, X. Ge, Y. Wang, and J. Fu, "Cloud Storage as the Infrastructure of Cloud Computing," in *Intelligent Computing and Cognitive Informatics (ICICCI), 2010 International Conference on*, 2010, pp. 380–383.

[31] Y. Huo, H. Wang, L. Hu, and H. Yang, "A Cloud Storage Architecture Model for Data-Intensive Applications," in *Computer and Management (CAMAN), 2011 International Conference on*, 2011, pp. 1–4.

[32] X. Guan and B.-Y. Choi, "Push or Pull?: Toward Optimal Content Delivery," in *Communications (ICC), 2011 IEEE International Conference on*, 2011, pp. 1–5.

[33] S. Borst, V. Gupta, and A. Walid, "Distributed Caching Algorithms for Content Distribution Networks," *INFOCOM, 2010 Proceedings IEEE*, pp. 1–9, 2010.

[34] S. D. Burd, G. Gaillard, E. Rooney, and A. F. Seazzu, "Virtual computing laboratories using vmware lab manager," in *Proceedings of the 2011 44th Hawaii International Conference on System Sciences*, ser. HICSS '11. Washington, DC, USA: IEEE Computer Society, 2011, pp. 1–9. [Online]. Available: http://dx.doi.org/10.1109/HICSS.2011.482

[35] P. Johnsen. Openload. [retrieved: April, 2012]. [Online]. Available: http://freecode.com/projects/openload