Proceedings of IOE Graduate Conference, 2016
pp. 35–39

Content Distribution and Localization: Sustainable and Efficient Model of Digital Internet Content Management

Krishna P Bhandari [1], Arun Timalsina [2]

[1] *Nepal Telecom*
[2] *Department of Electronics and Computer Engineering, Pulchowk Campus, Institute of Engineering, Tribhuvan University, Nepal*
Corresponding Email: [1] bhandari.kp@gmail.com

Abstract
Internet is essentially a network infrastructure that is used to carry digital data from one node to the other. Creation, storage, transfer and dissemination of that digital data is at the core of the Internet phenomena. With the rapid growth in the number and diversity of the services, variety of content, number of users and volume of data flow, it is becoming more and more challenging to properly store, transfer and manage the data. While it is important to create the content and reach it to consumers, it is equally important to make that access fast, reliable and user-friendly. The advent of rich content such as multimedia means it is not enough to just keep a high capacity link between the content source and the user. It is equally important to keep the content as close to the consumer as possible. It is a question of long term sustainability of the digital service and related content. This paper examines the major aspects related to local deployment and development of Internet content.

Keywords
Internet – Content – Content Delivery Network – Local Content – Caching – Proxy – Peering – Quality of Access

Content Access and Distribution

Content creation and management is an essential part of overall Internet experience. All sorts of services or applications over Internet need digital content at the back end. Traditionally the content consisted mainly of text and graphics. However, the recent trend is that the majority of content is audio/visual. The diversity of the content has also increased dramatically. Another important factor in quality of Internet experience is the proximity of the content. Though theoretically a digital content accessible over Internet can reside in any corner of the globe, in reality, the quality of access to the content dramatically differs based on the topological distance from the consumer. As a result, the intention of major content providers is to have their content present as closer to the potential audience as possible. This gives rise to special content distribution channels such as the CDN.

Content Management Stakeholders

The digital content ecosystem consists different stakeholders irrespective of the type of content. The principle stakeholders are the creator, the editor, the publisher, the administrator and the consumer. However, with the advent of social media, interactive web technologies and various file sharing and content upload services, a large number of people can be consumer and producer at the same time. This has also led to explosion in the amount of the content available online and the amount of content being created everyday.

One more important stakeholder is the ISP though it is not involved in actual creation and management of the principal Internet content. However, an ISP's entire business model revolves around providing its customers with easy, economical and reliable connection to the content.

Content Caching and Its Challenges

Traditionally, Internet content used to be largely static. This meant traditional caching techniques such as simple deployment of a proxy server yielded significant improvement in access quality and bandwidth saving. The idea was to keep a local, temporary copy of the content such as images, video clips etc. and serve the subsequent requests from the cache rather than from the original content source. This means saving of the bandwidth for the service provider and faster access for the consumer since the content is coming from local cache of the ISP. However, with advancement of technology, evolution of content and change in consumer habits, the content has become principally dynamic. As the content becomes more dynamic, traditional caching and sharing techniques become less and less effective. As a result, industry is moving away from traditional caching and towards collocation, peering and distribution. As illustrated in [1], the concept of content distribution through CDNs originated before 2000 and one can imagine what the Internet would be without them in the current era of rich content and billions of users.

Content Demographics

According to [2], Video and audio shall constitute 89% of the consumer Internet data traffic by 2018. This trend is generally universal. Since audio visual content is large in size, its consumption over the network needs a lot of bandwidth. This boom in audio visual content has also led to requirement of large bandwidth capacity to connect a network to the external world. This has affected the economy and business viability of a large number of service providers. At the same time, it has also posed a significant challenge to the content providers. They always fall short of the bandwidth required to pump content to ever growing number of consumers.

Explosion in number of devices and traffic

Mass proliferation of mobile telephony, increasing prevalence of data services in mobile telephony, advent of 3G and its successor technologies, introduction of IOT, among others, have led to remarkable increase in the number of online devices. It is estimated in [3] that the number of devices in a family of 4 shall dramatically rise in an average of 8 devices in 2012 to 24 devices in 2017 and 50 devices in 2022. The new devices shall mainly be automated, Internet connected IOT devices used inside the house and by the tenants. This ranges from automatic thermostats to IP based home appliances, entertainment devices, communication devices and others. All these devices are creating or consuming data in one way or another. This means need for more bandwidth, more storage and more processing power. This again necessitates promotion of local data storages and analytics to minimize pressure on the International and long-distance links.

This trend of the rapid increase in the number of connected devices according to the data published by Statista [4] and Cisco [5] is shown in Figure 1.

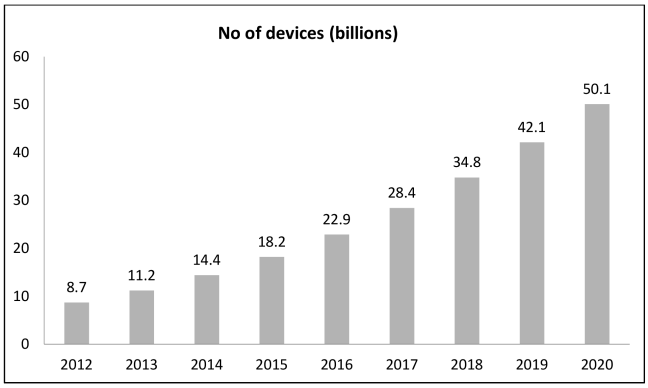

Figure 1: Number of Devices on the Internet

Impact of Cloud and IOT

The rapid advances in cloud computing and ushering in of the Internet of Things (IOT) era has also helped dramatically increase the number of devices creating or consuming digital data. The growing number of cloud services has also led to the Internet services being indispensable and the made the internet connectivity critical element of daily life. This means, higher need of ample and reliable bandwidth. This also necessitates the local deployment of as much content as possible. Nowadays, even cloud providers are shifting towards local deployment of content in terms of caches or private cloud deployments.

Introduction and predicted very rapid growth of the IOT phenomena also means unprecedented scenarios such

as entire smart cities, automatic sensors talking to each other, large amount of data moving to and from these sensors to the data storage/analytics platform [6]. This also means the development of local storage/analytics platform is critical for sustainable growth and successful operation.

Example Case Study

Deployment of the CDN, dedicated cache of a distributed content and even generic proxy or cache systems can become very important in making the internet service more sustainable for a service provider while improve the quality of experience for the subscribers or end users. A typical ISP in Nepal [7] has a yield of around 5% of its total traffic from a local Akamai cache server without any cost other than power for the machines and collocation. A 3-node Google Cache deployment has yielded more than 25Gbps (more than half of the total bandwidth consumed within the service provider) at a cost worth just a sixth of the IPLC capacity purchased to act as the cache feeding link to India and Singapore. This is more than 15-fold return on investment based on the current cost of international Internet bandwidth.

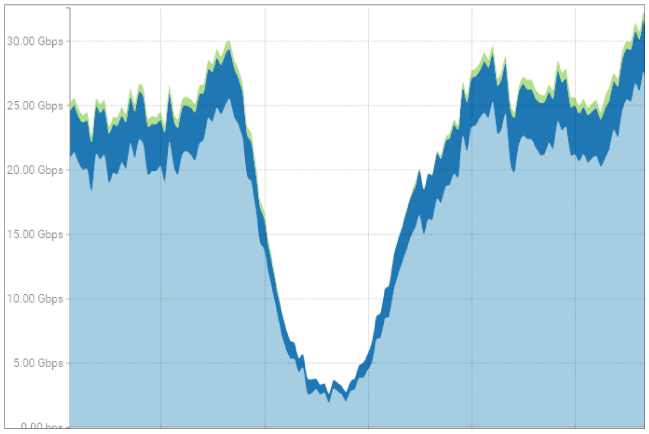

Figure 2: Traffic Profile of a Local Cache Deployment

A typical gain observed in the local-cache deployment of a major content provider is shown in Figure 2 [7]. There bottom portion is the total cache yield over a 20 hour period. The top portion is the total cache-feeding link traffic between the local cache and the overseas datacenter. The spillover traffic is shown in the narrow strip along the top of the graph. This shows that the ratio of total cache yield to the total overseas datacenter peering link is more than 6:1. This is a very beneficial model for the ISP as it provides high speed content while minimizing expenditure on the expensive international Internet bandwidth.

A further analysis of the same operator traffic highlighted the importance of fostering local content, by deploying caches, CDNs, local peering and other organic content development. The proportion of the traffic from different content sources is shown in Figure 3.

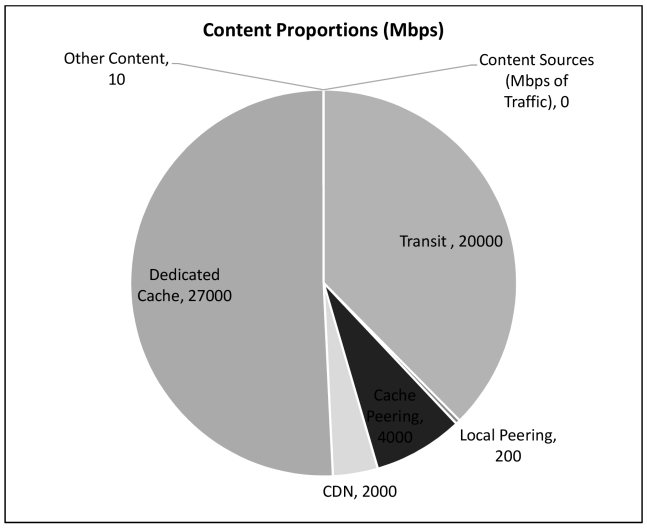

Figure 3: Traffic Distribution Among Different Sources

As illustrated in Figure 3, the network is benefitting heavily from the deployment of dedicated cache, CDN and direct peering. More than half of the total consumed traffic is from these sources. Only the rest is through the international Internet transit. As transit cost is the biggest investment of any large scale Internet service provider, such saving is extremely useful in making the services profitable for the operator while keeping them affordable for the subscribers. However, the data also indicates that there is a lot that can be done, especially in terms of developing organic content and local services.

Localization in the Interest of Content Owners

Deployment of the content nearer to the consumer is also in the interest of the owner of content. Large content such as Google, Facebook etc. use their own appliances or ride a CDN to deliver the content nearer to the consumer. A good quality of access for

consumers to the digital content owned by them is in their direct interest as their aim is to reach as many users as possible with good quality of access. Since they cannot build the entire network and ensure its quality, their strategy is to deploy their devices in the premises of service providers and large enterprises. This way, they can reach audience quickly without having to deploy the entire network. A report by OECD, ISOC and UNESCO highlights this in implying that content localization is critical for overall growth and enhancement of the digital economy [8]. The higher proportion of local content improves overall quality of Internet access, fostering local services and this has a domino effect of creating even more local services and content, progressively lowering dependence on international Internet bandwidth. This is a positive change cycle that builds iteratively on itself.

Moreover, the business model of CDN service providers like Akamai focuses on the outsourcing of the activity of reaching the content to the edge of the network. The content owners simply hand their content to the CDN operator who use their globally distributed network of distributed content appliances to deliver the content quickly to the end user without having to deploy exceptionally high capacity links to few large data centers [9]. They use thousands of nodes across the globe rather than a few very large ones. This technique, first used by the CDNs, are increasingly adopted by the other large content owners.

Such placement of content nearer to the users has marked impact on the round-trip latency between the user and the content. A September 2016 article in KeyCDN [10] shows that the deployment of CDN led to a more than 70% reduction in the average latency to a typical content from within the USA. Though the gain would not be as pronounced in lower-tiered networks such as in Nepal or other developing countries, the difference is still found to be significant. A typical latency to ISP-collocated and external Google content from an ADSL end user in Kathmandu is shown in Figure 1. It also shows that even the lowest latency to the external content is around 150ms whereas the local latency is steady around 50ms. This is a significant gain for the quality of access.

Localization in the Interest of Content Owners

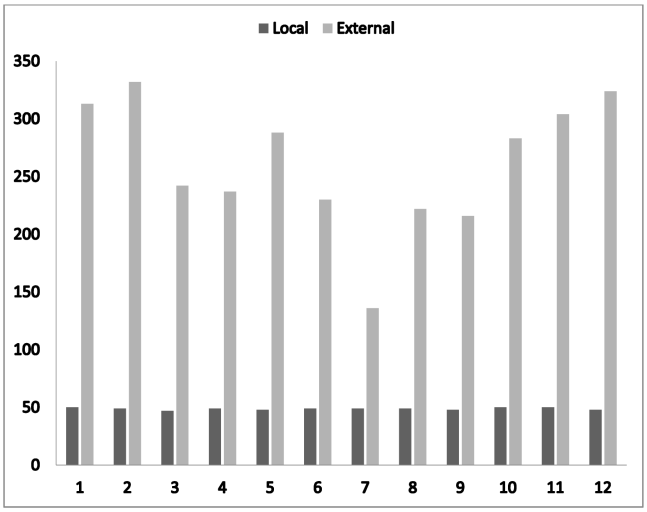

Figure 4: Sample Latency Comparison of Local and External Content

Summing Up

Burgeoning digital services, expanding coverage, increasing dependence of masses on the digital services, convergence of everything into the IP and online technologies, rapid increase in the volume of data consumed per person and overall growth of the digital economy indicates towards the sheer amount of content needed to drive all these growth trends. Sustainable Internet service depends upon the reachability of these contents and services to the masses in efficient and economical manner. Hence, it is of utmost importance to try to keep those content as close to the users as possible. This localization of the digital content can be driven by combination of a number of approaches including robust network, organic content, CDN, dedicated cache solutions and others. This is an ongoing and evolutionary process and needs constant change, optimization and dynamism on the part of the service providers as well as the content providers.

Future Work

This research can be further built upon to develop a content distribution and management model for different economies with different business models. Exact content management framework can have subtle differences among operators, enterprises and content providers in different economies and geographies. A single model can't be an exact fit for everyone.

References

[1] M. F. Kaashoek and F. Douglis. Scalable internet services. *IEEE Internet Computing*, 5(4):36–37, 2001.

[2] Deloitte. Digital media: The rise of on-demand content. Technical report, Deloitte, 2015.

[3] GSM Association. The impact of internet of things - digital home. Technical report, GSMA, 2015.

[4] Statista. Iot: Number of connected devices worldwide, 2016. URL: http://www.statista.com. [Accessed: September 17, 2016].

[5] Dave Evans. The internet of things: How the next evolution of the internet is changing everything. Technical report, Cisco, 2011.

[6] Jayavardhana Gubbia, Rajkumar Buyyab, Slaven Marusic, and Marimuthu Palaniswami. Internet of things (iot): A vision, architectural elements, and future directions. *Future Generation Computer Systems 29*, pages 1645–1660, 2013.

[7] Nepal Telecom. Nepal telecom - internet bandwidth status data, 2016.

[8] ISOC OECD UNESCO. The relationship between local content, internet development and access prices. *OECD Digital Economy Paper No.217*, 2013.

[9] John Dilley, Bruce Maggs, Jay Parikh, Harald Prokop, Ramesh Sitaraman, and Bill Weihl Akamai Technologies. Globally distributed content delivery. *IEEE Internet Computing*, pages 50–58, 2002.

[10] Brian Jackson. Website latency with and without a content delivery network, September; 21 2016. URL: https://www.keycdn.com/blog/website-latency/ [Accessed: September 23, 2016].

[11] Ernst & Young. Future of digital content consumption in india. Technical report, Ernst & Young, 2016.

[12] Vipul K. Gupta, Srinivasan Govindarajan, and Tonya Johnson. Overview of content management approaches and strategies. Technical report, Electronic Markets, 2001.