

# FUTURE OF PEER-TO-PEER TECHNOLOGY WITH THE RISE OF CLOUD COMPUTING

Phillip Kisembe<sup>1</sup> and Wilson Jeberson<sup>2</sup>

<sup>1</sup>All Nations University College, Koforidua, Eastern Region, Ghana

<sup>2</sup>Department of Computer Science & Engineering, SHUATS, Allahabad, India

## ***ABSTRACT***

*Peer-to-Peer (P2P) networking emerged as a disruptive business model displacing the server based networks within a point in time. P2P technologies are on the edge of becoming all-purpose in developing several applications for social networking. In the past seventeen years, research on P2P computing and systems has received enormous amount of attention in the areas of academia and the industry. P2P rose to triumphant profit-making systems in the internet. It represents the best incarnation of the end to end argument, the frequently disputed design philosophies that guided the design of the internet. The doubting factor then is why is research on P2P computing now fading from the spotlight and suffering a nose dive fall as dramatic as its rise to its popularity. This paper is going to capture a quick look at past results in peer-to-peer computing with focus on understanding what led to its rise, what contributed to its commercial success and what has led to its lack of interest. The insight of this paper introduces cloud computing as a paradigm to peer-to-peer computing.*

## ***KEYWORDS***

*Peer-to-peer networking, social networking, and cloud computing*

## **1. INTRODUCTION**

17 years ago research on P2P systems has received a significant amount of attention in the academic and industrial sectors. The main purpose of the P2P design is for peers to correspond on the internet, without the need of new protocols on switches and routers in the internet core. In P2P computing, nodes organize themselves as an overlay network, in which packet transmission on each of the overlay links uses standard Internet protocols, that is Transmission Control Protocol (TCP) and User Datagram Protocol (UDP). P2P file distribution systems allow nodes to collaborate among each another so that very bulky files can be distributed from one node to several receivers - gossiping. A file is divided into chunks, and nodes connect with one another in a random mesh topology, exchanging these chunks using random gossiping. Research on P2P systems began about 17 years ago and since its inception; it has gone through four major stages of evolution.

The earliest stage paid attention on well-organized search protocols, either by using a distributed hash table [1]-[3], or by using a gossip protocol. The second stage started around the year 2007, when the research attention on P2P systems shifted towards optimizing protocols to transfer large files. The third stage began around 2010 when live and on demand streaming protocols have become the topic that attracted passionate research attention. The final stage is where we are now which is P2P social networks [4]-[6]. These stages however have substantial overlaps in time. For example mechanisms that distribute resources in P2P have started to receive attention from cloud computing today.

P2P computing however rose into successful commercial systems, displaying the feasibility of their design. Despite all this, research in this field has recently declined and suffered a fall that we can equate to its rise in popularity [7]. Figure 1 represents the rise and fall in its popularity by counting the number of papers with the term P2P in its index. Data points are drawn from Advanced Google Scholar Search, with the term “P2P”

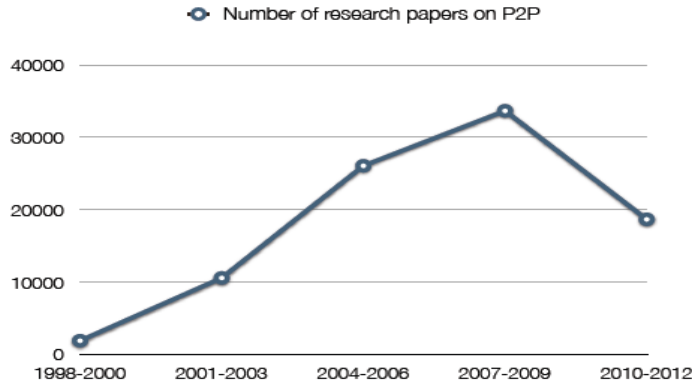


Figure 1. Rise and fall of the popularity of peer-to-peer as a research topic

Many can disagree of other ways of examining the fame of a research topic such as workshops, symposiums, and journals created to accommodate papers on the topic. When it comes to research in P2P computing, we can point out one workshop: the International Workshop on P2P systems (IPTPS). This workshop was launched by scholars from MIT in the year 2002, to offer a stage for researchers involved in this field to discuss various challenges. It became an annual event and attendance was strictly by invitation. Yet in 2015 according to [8], the annual workshop was cancelled with emphasis being put on lack of interest in the subject matter.

Why then do we observe a rise and an equally drastic fall of research interest in P2P systems? This paper attempts to offer a review of the past 17 years of P2P research and show what contributed to its rise in popularity in academia, the key features that led to its commercial success, and what has led to its research fall. Observations made in this paper may be useful to shed some light on our thoughts on cloud computing research.

## 2. EVOLUTION OF PEER-TO-PEER FILE SHARING SYSTEMS

The most important cutting edge applications that motivated peer-to-peer system design is file sharing. Users very often need to distribute their files to others as conveniently and as efficiently as possible for example friends on a social group. These files are essentially large in their sizes, such as a HD song, or a video, and large PDFs in recent years. [9] Proposes two fundamental issues that arise from such a need of sharing large files: first need to have an effective way to search for their availability, and then an efficient mechanism to distribute the file from its source to all the users who may be interested.

The P2P file sharing applications can be categorized into two types: centralized/hybrid or decentralize/pure architectures [10]. In the centralized model, for example Napster and Bit Torrent, there is the presence of a central server to index contents and peers. This model is easy to manage and gives a competent search. However, it has a limited scalability and a single point of failure. Decentralize p2p systems eliminate a central server, thus enhancing the reliability [11].

Peer to Peer, previously known as the host-to-host concept, was at first used to institute equal communication between two peers. In this concept each peer plays the role as server and client simultaneously. USENET was established for exchanging information between UNIX machines.

### 2.1. NAPSTER

Napster is a well known P2P file sharing application; written by "Shawn Fanning" in 1999. Napster was first created for those who had difficulty in discovering and downloading MP3 music files. Napster is a combination of a search engine and a file-sharing application. Napster is a file sharing service that facilitates the location and exchange of files, which usually include images, audio or video, via access to the internet. Napster being the first sharing application paved way for various programs including Kazaa, Morpheus, LimeWire and Bearshare. Figure 2 shows the architectural structure of Napster. However, in 2001, [12] Napster was penalized by the Recording Industry Association of America (RIAA) for copyright reasons and was forced to pay a fee of \$26 million to resolve its legal battle with songwriters and music publishers. Figure 2 below depicts an architectural representation of Napster and how messaging occurs between a directory server and the connected users.

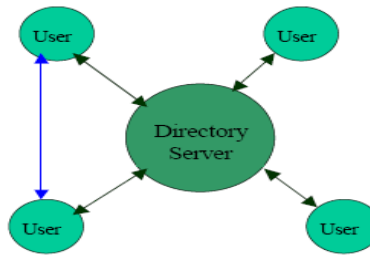


Figure 2: Architecture of Napster

### 2.2. GNUTELLA

Gnutella is a large P2P network. Other networks adapted this model since it was first of its kind. Peers have same capability and responsibility. Nodes communicate in a symmetric manner in this type of network. Lack of a central directory server Index allows metadata of shared files to be stored locally among all peers. Figure 3 below represents the architectural representation of Gnutella in which: Each user acts independently. Nodes communicate with each other with the use of queries.

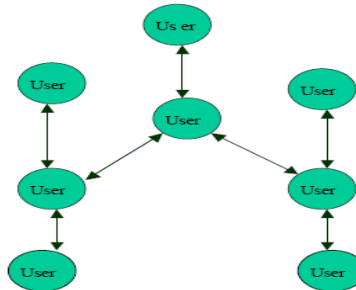


Figure 3: Architecture of Gnutella

### **2.3. KAZAA**

KaZaA uses P2P file sharing; this is a similar technology which led to the rise of Napster. When compared to Napster, that distributed files through a central server, KaZaA uses a decentralized system. KaZaA users make contact with one another directly online to share resources. To transfer data between peers, KaZaA uses the FastTrack protocol.

FastTrack is a so-called "second generation" P2P protocol [13]. The system divides KazaA users into two groups: supernodes and ordinary nodes. Supernodes are mainly computers with high processing capabilities.

Each supernode can manage 60 to 150 ordinary nodes at any particular time. Users have to install the KaZaA software on the computer, which already has a list of supernodes. Every time this application is launched by the user, the computer registers with the central server and then selects from a list of currently active supernodes. Once a file request is sent out by the computer to either upload or download a file, this request will be through the supernode.

The supernode communicates with other supernodes which later on connects with other regular nodes in order to fulfill the request until the Time to Live of 7 runs out. This indicates that the search request will extend seven levels into the network before it stops propagating. When the file needed is located, it is transferred directly from the source to the destination using HTTP (HyperText Transfer Protocol).

### **2.4. BIT-TORRENT**

Bit-Torrent appeared in 2005 as an application layer network protocol used to distribute files. Nodes here can download and upload a file at the same time. This does not necessary mean that a node finishes its download before uploading. Complete downloaded blocks can be uploaded to other nodes on the network.

The serving capacity increases as the number of downloads increases making the system self-scaling [14]. Client-server architecture is also adapted here as nodes contact the server to locate other nodes that they may connect to. A torrent file must first be acquired for one to begin a download [15]. The .torrent files are found on the web and contain information about; the files that are contained in the torrent the location address/hostname of the torrent's tracker The tracker is a computer which acts as a server, and maintains a list of nodes which are participating in the system (i.e. computers that are currently downloading or uploading file data) [16].

Once a computer has established a connection to one other node in the system, it can begin downloading data. Faster download occurs based on the increased number of node connections hence the likelihood of obtaining a complete copy of the file. A typical Bit-Torrent system needs to have at least one seeder which has a complete copy of the data being shared.

When a copy has been uploaded by the seeder to other nodes, the seeder may leave the network. However, for successful ongoing sharing of the data, at least one complete copy must remain available across the nodes in the network.

### **2.5. PEER-TO-PEER VIDEO STREAMING**

In 2009, video streaming of content on the internet became popular due to the increase in demand of multimedia content. This led to the modification of network technologies.

A broad collection of such applications can be found in these days, e.g. as video streaming, video conferencing, surveillance, media broadcasting, e-learning and storage, etc. In regular, video streaming techniques typically employ the client- server model and the usage of Content Distribution Networks (CDN) along with IP multicast which has been the most desirable solutions to support media streaming over the Internet.

Each node contributes with its own resources to the streaming session [18 – 20]. This serves as an advantage in using P2P content delivery. For high quality service to be effectively provided in this network, P2P protocols need to be modified so that they are able to support multimedia features and video streaming technology needs in order to cope with the changing bandwidth capacities. [21] Some of the common P2P based video streaming platforms include; PP Live – that is widely used in China, Vuze – a P2P based video on demand platform, and BBC player.

## **2.6. DISTRIBUTED SOCIAL NETWORKS**

Distributed social networks became popular in 2014 and are currently picking up today. According to [22] P2P file sharing system provides a platform for a great number of nodes to share their files. They first study trace data from Facebook and BitTorrent. Guided by the observation a system is proposed that integrates social network into a P2P network named social P2P, for simultaneous efficient and trust worthy file sharing. Three mechanisms are incorporated:

- a) Interest based structure
- b) Interest based file searching
- c) Trust relationship adjustment

By exploiting the social interest and relationships in the social network, the interest based structure groups' common-multi-interest interest nodes into a cluster and further connects socially close nodes within a cluster. The comparably stable nodes in each cluster form a Distributed Hash Table (DHT) for inter-cluster file searching. In the interest based file searching mechanism, a file query is forwarded to the cluster of the file by the DHT routing first. Then, it is forwarded along constructed connections within a cluster, which achieves high hit rate and reliable routing. Sharing files among socially close friends discouraged nodes from providing faulty files because people are unlikely to risk their reputation in the real world. In the trust relationship adjustment mechanism, each node in a routing path adaptively decreases its trust on the node that has forwarded a faulty file in order to avoid routing queries towards misbehaving nodes later on. Results show that social P2P achieves highly efficient and trust worthy file sharing compared to current file sharing systems and trust management systems.

## **3. COMPARISON OF PEER-TO-PEER TECHNOLOGY TO OTHER TECHNOLOGIES**

In this section the similarities and differences of the client/server and cloud computing models are discussed in relation to the P2P technology. Both the client/server and cloud computing models are successful technologies and are the underlying architecture of the internet today too, but what made P2P popular and different? This section shows what makes a P2P technology special as compared to the other mentioned technologies.

### **3.1. CLIENT/SERVER MODEL VS. P2P TECHNOLOGY**

Seventeen years ago P2P technology was not in existence but a server which clients connected to get the information that they wanted. Figure 4 shows the architecture of the client/server model. The major challenge here was that the clients did not know precisely on which server the information they wanted was resting on. This client was then later to be directed to another server

which is the technology that led to the rise of search engines such as Google and yahoo which do the searching of the server for the client. It is easy to set up if you don't have very many users and also if something goes wrong you can easily fix it hence making it very reliable and with well known behaviour but if the server crashes it will not provide enough memory to serve all the clients leading to resources not being available. The client/server model is clearly not P2P since communication is only between clients and servers, not between clients and clients.

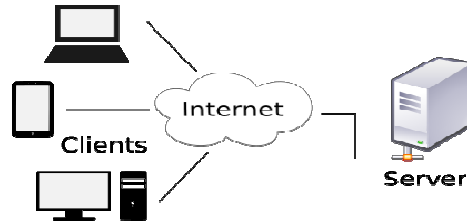


Figure 4. Client/Server Architecture

### 3.2 CLOUD COMPUTING VS. P2P TECHNOLOGY

In cloud computing we have a distributed pool of resources, basically it's not just one server but a federation of several servers. For example face book is estimated to have over 100,000 servers running their infrastructure so is Google. But despite this every node can't fail and they all provide additional benefits such as storage and computational power and typically they are all operated in a controlled environment. Therefore, we don't have malicious nodes, they don't go offline and if they do they are just thrown away at some point and operate on homogeneous devices. Cloud computing also has selective centralized structures with single access point where centralized updates can be performed and monitoring. Figure 5 below shows the architecture of a cloud computing environment.

Some of the essential cloud computing aspects it provides are on demand self-service where resources such as server time and network storage are automatically provided to a customer when required. Second being rapid elasticity, the underlying infrastructure is able to adapt to changing requirements e.g. number of concurrent users. It also allows for dynamic up/down scaling. Third is measured service which is the metering of resource and service consumption to provide elastic pricing and billing models e.g. pay per use. Fourth is resource pooling that provides and assigns dynamism in a multi-tenant way. Finally, broad network access capabilities are available worldwide over standard network mechanisms.



Figure 5: Cloud Computing Architecture

### **3.3 NEED FOR P2P SYSTEMS**

These systems are very similar to cloud computing since they do have a large pool of resources, but it is in an uncontrolled environment therefore churning is performed, there is the availability of heterogeneous devices with different capabilities, they are unpredictable with distributed access points. Currently the mechanisms that do this distributed data and search in cloud computing are those used in the P2P environment. The question then is why not use cloud computing since it provides all the nice features? Why the need for peer to peer systems?

Several reasons are available to answer this one being P2P protocols serve as a backend to cloud computing technology. Decentralized mechanisms, data storage and big data are provided by the P2P protocols. Various decentralized mechanisms, data storage mechanisms and the massive data available on the servers in the cloud operate on P2P protocols in the sharing of resources in the cloud. Two is costs; resources are available and are for free. E.g. Facebook has over 100,000 servers for more than a billion uses. So how much does it cost to operate a year per user? It is estimated to 2 dollars per year for every user [17]. Third being security which has to be integrated in all the mechanisms. There should be no central point to get data over all users. Protocols must be designed to assume that everyone is an attacker. Fourth is the locality/"offline" usage where local data is exchanged in companies or private households. The fifth point is a larger set of resources are available. Some of the resources available can be categorized as content which include programs, data and information and also various capacities such as CPUs, storage, memory and bandwidth. Finally observation, the quality demand of users of the content grows basing on two things; the processing speed and amount/size of content. Demand goes beyond the capabilities of single devices.

### **4. MAJOR CHALLENGE OF P2P**

The rise of research interests in peer-to-peer computing was largely fuelled by real-world peer-to-peer systems from the industry, most notably Gnutella, as well as by the need of millions of users to share large files in a peer-to-peer fashion. The essence of peer-to-peer systems is to take full advantage of both storage and bandwidth resources on end hosts, realizing potentially substantial resource savings on dedicated servers. For example, peer-to-peer live streaming systems were designed to utilize the upload bandwidth of end hosts to alleviate bandwidth demand on servers in Content Distribution Networks (CDNs). Peer-to-peer file sharing, on the other hand, utilized peer upload bandwidth to improve the downloading performance.

The other major contributing factor to the strong academic interests, in our opinion, was the fact that the design of peer-to-peer systems was able to start from a clean slate, in that it was not confined by any legacy protocols in the Internet.

Major drawbacks of peer-to-peer systems started to be realized, with extensive research efforts to solve them. The number one challenge on the list was its lack of robustness against peer dynamics, such as a flash crowd scenario in which a large number of peers join around the same time, or an alarmingly high attrition rate in which a large number of peers leave the system. By dividing the file into smaller blocks and serving them from multiple peers, BitTorrent was able to mitigate the negative impact of peer dynamics, but was still not able to provide any performance guarantees that are routine if dedicated servers are to be used. This still leaves the big challenge of free-riders in P2P content distribution. Various incentive mechanisms have been proposed to curb this problem such as tit for tat mechanism that is employed by bit torrent to minimize the free riding issue. In this case nodes that upload a file are given an incentive of an increase in their bandwidth. This solution however, has still not proved to be effective to this challenge.

## **5. PROMISE OF P2P CLOUD COMPUTING**

Many of the most popular and successful Internet services today such as Netflix and Dropbox make use of cloud computing. Cloud based data centres are costly to construct and are mostly owned by huge organizations such as Google and Microsoft. Recent research has been investigating a very different strategy to do cloud computing without those giant centralized facilities at all—using peer-to-peer technologies of the kind sometimes associated with shady file-sharing operations.

Placing the physical infrastructure for a cloud-computing operation where it's usually found, in a single massive data centre, has definite advantages. Construction, equipment procurement, installation, and maintenance are all simplified, and economies of scale reduce costs. On the other hand, a single large data centre consumes an enormous amount of electrical power, often comparable to what you'd need to run a small town, and dissipating the waste heat it generates is usually a big headache.

Another drawback of centralized clouds is that their geographic location, which may be best for the owners, may not be best for the customers. This is the case, for example, when governments place restrictions on sensitive data crossing national borders. A data centre located in one country may then be off-limits to customers in some other countries.

In principle, a P2P cloud could be built using the ordinary computing, storage, and communication equipment found now in people's homes, with essentially zero initial investment. Broadband modems, routers, set-top boxes, game consoles, and laptop and desktop PCs could all contribute. The challenge is to turn this motley collection into a coherent and usable cloud infrastructure and offer its services to customers. You also have to ensure that the salient features of clouds—on-demand resource provisioning and the metering of service—are maintained.

This would surely be tough to do, but think of the advantages. First, there would be no single entity that owns or controls it. As with most other P2P applications, a P2P cloud could be created and operated as a grassroots effort, without requiring the permission or consent of any authority. People would choose to participate in a P2P cloud by installing the appropriate client software on their local machines, and the value of the resulting cloud infrastructure would be commensurate with the number of individuals who are contributing to it.

A second advantage comes from the fact that a P2P cloud's components are small, individually consume little power, and well distributed. This drastically reduces concerns about local catastrophes. It also removes the usual worries about heat dissipation. Although such P2P clouds couldn't provide the quality-of-service guarantees of a Google or an Amazon, for many applications that wouldn't much matter.

## **6. CONCLUSION**

As cloud computing begins its rise in taking over as the current technology, can P2P be rescued from its fall in popularity? This paper has covered a journey of the P2P technology and you will notice that from file sharing and video streaming to social P2P, this technology overlaps through time. File sharing is still being implemented in social P2P. This is because, the social groups that are coming up on the internet today that are defining which functionality is current on the internet are youths and adults using snapchat, whatsapp, Facebook, etc. User generated content is becoming more and more popular today. Still the shift in paradigm to cloud computing is unstoppable but yet cloud computing implements P2P file sharing protocols in the sharing of its resources in the cloud. The success of the many volunteer-computing projects demonstrates the



extreme scale that a P2P cloud could in principle attain, both in terms of the number of different computers involved and their geographic distribution. Using such a collection would, of course, mean that equipment failures will be common. And besides, the people who contribute their computers to these clouds could turn them on and off at any time, something the people who run P2P networks refer to as “churn.” So the first task for any P2P cloud is to keep track of all functioning and online devices enrolled in the system and to dynamically partition these resources among customers. And you have to do that in a completely decentralized manner and despite churn. In summary P2P computing is a technology still in use today and should not be lost.

## ACKNOWLEDGEMENTS

The authors would like to thank everyone who has contributed to the completion of this paper.

## REFERENCES

- [1] E. K. Lua, J. Crowcroft, M. Pias, R. Sharma, and S. Lim, (2005) “A survey and comparison of peer-to-peer overlay network schemes”, *IEEE Communications Surveys and Tutorials*, 7(2):72–93. <http://www.disi.unitn.it/~montreso/ds/papers/P2PSurvey.pdf>
- [2] Stoica, R. Morris, D. Karger, M. F. Kaashoek, and H. Balakrishnan, (2001) “Chord: A Scalable Peer-to-Peer Lookup Service for Internet Applications,” in *Proc. ACM SIGCOMM*, pp. 149–160.
- [3] Rajput, Ubaidullah, Fizza Abbas, Jian Wang, Hasoo Eun, and Heekuck Oh, "CACPPA: A Cloud-Assisted Conditional Privacy Preserving Authentication Protocol for VANET", In *Cluster, Cloud and Grid Computing (CCGrid)*, pp. 434-442, 2016.
- [4] Peer-to-Peer Systems, (2001)” in *Proc. International Conference on Distributed Systems Platforms (Middleware)*, pp. 329–350.
- [5] D. Frey, A.-M. Kermarrec, M. Raynal, J. Stainer, (2013) "Trust-aware peer sampling: Performance and privacy tradeoffs", *Theor. Comput. Sci.*, vol. 512, pp. 67-83.
- [6] W. Galuba, K. Aberer, Z. Despotovic, W. Kellerer, (2010) "Leveraging social networks for increased bittorrent robustness", *Proc. 7th IEEE Conf. Consumer Commun. Netw. Conf.*, pp. 159-163.
- [7] H. Shen, Ze Li, and K. Chen, (2015) “Social P2P: An online social network based P2P file sharing system”, *IEEE Transactions on Parallel and Distributed Systems*, Pg. 2874-2889.
- [8] R. Yu et al., "Cooperative Resource Management in Cloud-Enabled Vehicular Networks," in *IEEE Transactions on Industrial Electronics*, Vol. 62, No. 12, pp. 7938-7951, 2015.
- [9] Mozilla Development Network (MDN). Retrieved April 10, 2017 from <http://www.wikicfp.com/cfp/program?id=1651&s=IPTPS&f=International%20Workshop%20on%20Peer-to-Peer%20Systems>
- [10] G. Dan, N. Carlsson, and I. Chatzidrosso, (2011) “Effective and highly available peer discovery: A case for independent trackers and gossiping”, In *peer-to-peer (P2P) computing, 2011 IEEE International conference on*, pages 290—299.
- [11] Clarke, L.; Sandberg, O.; Wiley, B. Hong, T.W, (2009) “Freenet: a distributed anonymous information storage and retrieval system”, *Design Privacy Enhancing Technologies (Lecture Notes in Computer Science*, vol. 2009), p. 46 – 66. *International Workshop on Design Issues in Anonymity and Unobservability*, Berkeley, CA, USA *Proceedings*
- [12] Lu, Yanrong, Lixiang Li, Haipeng Peng, and Yixian Yang, "Robust ID based mutual authentication and key agreement scheme preserving user anonymity in mobile networks", *KSII Transactions on Internet & Information Systems*, Vol.10, No.3, 2016.
- [13] Mozilla Development Network (MDN). Retrieved April 10, 2017 from <https://www.wired.com/2009/12/1207riaa-sues-napster/>
- [14] Mozilla Development Network (MDN). Retrieved April 10, 2017 from <http://computer.howstuffworks.com/kazaa3.htm>
- [15] "RIAA v. The People: Five Years Later". *Electronic Frontier Foundation*. Retrieved 1 November 2016
- [16] Bharambe, Ashwin R, (2003) “Analyzing and Improving BitTorrent Performance,” in *MSR-TR*
- [17] Cohen, Bram, (2003) “Incentives Build Robustness in BitTorrent”.
- [18] Mozilla Development Network (MDN). Retrieved April 10, 2017 from <http://www.bittorrent.com/bittorrentecon.pdf>

- [19] M. Castro, P. Druschel, A.-M. Kermarrec, A. Nandi, A. Rowstron, and A. Singh, (2013) “Splitstream: High-bandwidth content distribution in cooperative environments”, In Proceedings of IPTPS03, Berkeley, USA.
- [20] Rachel King, (2014) “Facebook data centers worldwide, by the numbers and pictures”.
- [21] X. Zhang, J. Liu, B. Li, and Y.-S. Yum, (2005) “Cool streaming/donet: a data-driven overlay network for peer-to-peer live media streaming,” in Proc. IEEE INFOCOM, vol. 3, pp. 2102–2111.
- [22] Y. Huang, T. Z. Fu, D.-M. Chiu, J. C. Lui, and C. Huang, (2008) “Challenges, Design and Analysis of a Large-scale P2P-VoD System,” in Proc. ACM SIGCOMM, pp. 375–388.
- [23] C. Huang, J. Li, and K. W. Ross, (2007) “Peer-Assisted VoD: Making Internet Video Distribution Cheap,” in Proc. 6th International Workshop on Peer-to-Peer Systems (IPTPS), pp. 1-6.
- [24] N. Ramzam, H. Park, and E. Izquierdo, (2012) “Video streaming over P2P networks: Challenges and Opportunities”. Elsevier.
- [25] J. Ross, Y. Lin, and S. K. Chalup, (2016) “RTCSS: a framework for developing real time peer-to-peer web application”, ACSW ‘16’ proceedings of the Australian Computer Science week Multi-conference, Article No.56

## AUTHORS

**Phillip Kisembe** is currently a Ph.D candidate doing his research on P2P systems at SHUATS, India. He is under sponsorship from All Nations University College, Ghana.



**Wilson Jeberson** is currently the Head of Department of Computer Science and Information Technology at SHUATS, India.

