PSHARE: A PREDICTIVE VIDEO SHARING STRATEGY IN PEER-ASSISTED CONTENT DELIVERY NETWORKS

CHAPTER ONE

INTRODUCTION

1.0 BACKGROUND OF THE STUDY

A computer network allows computers or other computing devices to exchange data or resources, devices in a network share resources through communication links (Ahlawat and Anand, 2014). Communication links between nodes are established using either cable media or wireless media. The Internet is the largest network ever created by mankind. The Internet connects thousands of millions of computing devices including Personal Computers, Laptops, Workstations, Server, Smartphones, tablets, Televisions, Webcams, Environmental devices, Automobiles, Security Cameras and many more via different communication links such as wireless radio signals, fiber optical cables, twisted pair cables, untwisted pair cables and coaxial cables. Networking devices that originate, route and terminate the data are called network nodes (Ahlawat and Anand, 2014). In today's communication market, many types of networks exist but the most distinct types of networks are Local Area Networks and Wide Area Networks (Furth and Hari, 2015). In a LAN, computers are connected within a “local area” (for example, an office or home). In a WAN on the other hand, computers are farther apart and are connected via telephone/communication lines, radio waves or other means of communication.

Networks can be classified based on topology, protocols and architectures. Network topology refers to the way in which the network of computers is connected, it basically means the technology for arrangement of various computer elements like communication links, nodes and so on (Anjum and Pasha, 2015). Network topologies are categorized into the following basic types: point-to-point, bus, ring, star, tree, mesh and hybrid topologies. A point to point topology is a direct connection between two devices (nodes), it is the simplest of all the network topologies. Point-to-point links can be Simplex, Half-duplex or Full-duplex. In Simplex signal flows in one direction and only one station transmits and the other receive, in Half-duplex each station can both transmit and receive but not at the same time and where as in Full-duplex both stations transmit and receive simultaneously (Anjum and Pasha, 2015). The Bus topology uses one main cable to which all nodes are directly connected, one of the computers in the network typically acts as a server and the main cable acts as the backbone of the network. In Star topology, each computer is connected to a central hub using a point-to-point connection, the central hub can be a computer server that manages the network, or it can be a much simpler device that only makes the connections between computers over the network possible. The major advantage of the star topology is that if one connection fails, it does not affect others (Anjum and Pasha, 2015). In ring topology, the computers in the network are connected in a circular fashion, and the data travels in one direction, each computer is directly connected to the next computer, forming a single pathway for signals through the network. In mesh topology, every node has a direct point-to-point connection to every other node, because all connections are direct, the network can handle very high-volume traffic, it is also robust because if one connection fails, the others remain intact. Tree topology combines multiple star topologies onto a bus, hub devices for each star topology are connected to the bus, each hub is like the root of a tree of devices, this provides great flexibility for expanding and modifying the network. A hybrid topology uses a combination of two or more topologies. Hybrid networks provide a lot of flexibility, and thus, they have become the most widely used type of topology.

Network Architecture is basically the physical and logical design of a network, it refers to how computers are organized and how tasks are allocated among these computers. the client-to-server (CS) networks, the peer-to-peer (P2P) networks and the content delivery networks (CDN) are types of Network Architectures.

A client-to-server network can be defined as a software architecture made up of both the client and server, whereby the clients always send request while the server responds to the request sent (Haroon Shakirat Oluwatosin, 2014). Benefits of the CS network includes it splits the processing of application across multiple machines, it allows sharing of resources from clients to servers, it reduces data replication by storing data on each server instead of client, it is not restricted to a small number of computers and the server can be accessed anywhere and across multiple platforms (Haroon Shakirat Oluwatosin, 2014).

A Content Delivery Network is a variation of the client-server service network (Yong Liu *et al,* 2017). A CDN is a system of distributed servers that deliver videos and other web contents based on the geographical location of the user, the origin of the web page and the content delivery server, It is a two-layer client-to-server network. The term CDN implies a network infrastructure that supports the distribution of content (Plagemann T et al, 2015). Content in this context consists of encoded data or multimedia data for example video, audio, documents, images, web pages and metadata that is data about data. Content can be pre-recorded or retrieved from live sources (Plagemann T et al, 2015). It stores a cached version of its content in multiple geographical locations. This service is effective in speeding the delivery of content of websites with high traffic and websites that have global reach. The closer the CDN server is to the user geographically, the faster the content will be delivered to the user. Advantages of using a content delivery networks are: lower server load, faster content delivery to the web browsers, increase in concurrent users on a server and customer may have already downloaded the content. The main disadvantages of using a content delivery network are: content delivery networks cost additional money, adds complexity to your website, customers may have network filters that block some Content Delivery Networks and prevent your content from being uploaded and the Geo-location may not be close to your target audience.

In Peer-to-peer(P2P) Networks each node is equal in terms of functionality (Mourad Amad *et al,* 2012). A Peer-To-Peer network consists of a group of computers each of which acts as both clients and servers and contribute their resources, the network is configured to allow certain files and folders to be shared with everyone or with selected users. The major advantages of a peer-to-peer network is that it does not require a dedicated server which makes it less expensive, if one computer stops working, the other computers connected to the network will continue working, Installation and Setup is quite painless because of the built-in support in modern operating systems. The disadvantages of the network are security and data backups are to be done to each individual computer and as the numbers of computers increases on a P2P network performance, security and access becomes a major headache.

Peer Assisted Content Delivery Networks (PA-CDN) is a hybrid architecture which aims to combine the advantages of both Content Delivery Networks (CDNs) and Peer-to-Peer (P2P) systems (Nasreen Anjum *et al*, 2017). The CDN servers act as a backup node in the P2P content distribution. PA-CDN allows peers to distribute chunks of contents amongst each other whenever there is sufficient capacity in the swarm, however when there is no sufficient capacity in the swarm (for example, there are no peers near a user with free upload to deliver the content) users are served directly from CDN servers (Nasreen Anjum *et al*, 2017). In the PA-CDN, usually the first user collects data from the server and then acts as a server to other users. It reduces the drawbacks of pure P2P systems by replicating content in a minimal amount of edge servers, therefore enabling reliable content distribution with a guaranteed quality of service (Nasreen Anjum *et al*, 2017).

Multimedia is a technology that enables humans to use computers capable of processing textual data, audio and video, still pictures and animation (Nankya Mariam et all, 2017). It allows computer users to communicate with the computer system in a variety of ways (speaking, writing, moving objects and so on), data is acceptable in the form of voice, moving pictures, images and videos and does not have to be converted into machine readable format. Multimedia can be used in schools to help boost students learning process, it can also be used to give a real world impression while using the computer. Video and Audio are the most popular multimedia application. Digital audio has lower bandwidth compared to video although users tend are more sensitive to audio malfunctions than video. A Video has high bit rate (100Kbps for low-quality video conferencing to 3Mps HD – High Definition movies). Videos can be compressed, compression can be used to create different versions of a vide with each having different video quality (Nankya Mariam et all, 2017).

Since the very early stages of the Internet, it has always been the belief that videos should be streamed to the users over the network: a user can begin playing a video segment before the entire video has been transmitted (Baochun Li *et al,* 2015). Video streaming is a type of media streaming in which the data from a video file is continuously delivered via the Internet to a remote user. It allows a video to be viewed online without being downloaded on a host computer or device. An estimated one-third, of all online activities on the internet is spent watching videos (Nasreen Anjum *et al*, 2017). Video streaming can be transmitted via different network architectures such as peer-to-peer networks, content delivery networks and peer assisted content delivery networks. In the content delivery network, the video source server first push video content to a set of content delivery servers placed strategically at the network edges Network (Yong Liu *et al,* 2008). CDN effectively shortens the users' startup delays, reduces the traffic imposed on the network, and serves more users as a whole. YouTube employs CDN to stream video to end users (Yong Liu *et* *al,* 2008). The major challenge for server-based video streaming solutions, is its scalability. In the peer-to-peer network, video streaming takes the advantage of the ability of participating end hosts, or peers, in a multicast group to contribute their uplink bandwidth (Baochun Li *et al,* 2015). The peer-to-peer network has two major advantages. First, it does not require the support from the underlying network infrastructure, and as a result, it is cost-effective and easy to be deployed. Second, in the peer-to-peer design, a peer is not only downloading a video stream, but also uploading it to other peers watching the same program. In Peer-Assisted Content Delivery Streaming, the initial peer downloads chunks of video from the Content Delivery Servers and serves as a server to other peers. Other peers then logging into the system and requesting the same content download from the initial peer and then act as a server to other peers too. This research focuses on improving video streaming in PA-CDNs by using prediction.

1.2 MOTIVATION

Video streaming has been an increasingly popular application in the Internet. Facebook is now the second-largest online video viewing platform. The total time spent on video viewing on Facebook increased 1840 percent year-over-year, from 34.9 million minutes in October 2008 to 677.0 million minutes in October 2009 (Haiying Shen *et al*, 2014). Netflix alone is reportedly streaming over 1 billion videos each month and the figure is rising constantly (Nasreen Anjum *et al*, 2017).

The increasing demand for video traffic have questioned the effectiveness of traditional Content Delivery Networks. The proposed scheme will deploy the hybrid network (PA-CDNs) to enable effective video transmission from one user to another, from one geographical location to another.

1.3 STATEMENT OF PROBLEM

Delegating content delivery functions to peers is criticized because of the peers unpredictable and unreliable nature (Nasreen Anjum *et al*, 2017). P2P systems suffer from various problems like partial participation - users partially participate when they abandon a video session after watching the first few chunks, heterogeneity in resources - different users have different Internet connections and different resource capabilities both affecting how much and how efficiently contents can be shared with other peers which affects quality of service. In most Social Networks (SN), most of the viewers of a user's videos are the user's close friends, most video views are driven by social relationships, and the rest are driven by interest, and viewers of the same video tend to reside in the same location (Haiying Shen *et al*, 2013). This is called the Friend-to Friend Network which is a type of P2P network in which users only make direct connections with people they know). Users generally have preference for some unique video contents ranging from scientific videos to music videos to short clips of videos (ranging from fifteen seconds to two or three minutes) to long length videos. Users preferences are gradually learnt from their historical video access patterns User video preference sometimes determine how long a user will be willing to spend in the system, if the user will be willing to share upload bandwidth and if the user will be willing to allow other users to have access to it. User video preference can be optimized to reduce startup delay and playback latency with prediction, initial chunks will be downloaded from CDN nodes and delegating the remaining work to peers. To optimize user video preference, when a user logs into a video content provider, the system can predict the video a user will be willing to watch using user video preference. Thus, the system downloads chunks of the video from other peers faster and can also delegate the user to the Content Delivery Server if need be.

1.4 AIMS AND OBJECTIVES

The aim of this research is to develop an improved video sharing predictive scheme suitable for peer assisted content delivery networks.

The specific objectives are to:

a. Review existing literature in multimedia applications, Network architectures and NeuroFuzzy.
b. Formulate a predictive video sharing model in Peer Assisted Networks.
c. Simulate the Model in (b) using MATLAB simulation tool and Network Simulator 3.
d. Evaluate the model.

1.5 METHODOLOGY OVERVIEW

1.6 ORGANIZATION OF WORK

The project work contains five chapters and it is organized as follows:

Chapter one introduces the title of the project. It gives the background of the study, motivation, the problem statement, aims, objectives and methodology overview.

Chapter two involves literature review. This chapter also includes the review of journals related to the project work.

Chapter three presents the methodology of the study.

Chapter four is about the implementation of the system.

Chapter five presents the conclusion of this research work.

CHAPTER TWO

LITERATURE REVIEW

2.1 MULTIMEDIA

Multimedia can be described as a way of handling a variety of representation media in an integrated manner. Multimedia is concerned with the computer-controlled integration of text, graphics, drawings, still and moving images (Video), animation, audio, and any other media where every type of information can be represented, stored, transmitted and processed digitally. This means that the media sources are integrated into a single framework (Govind *et al.*, 2010).

A Multimedia Application is an application which uses a collection of multiple media sources e.g. videos, audios, text, graphics and so on. Examples of Multimedia Applications includes the world wide web, video conferencing, video-on-demand, interactive television, home shopping, games, virtual reality, digital video editing and production systems, multimedia database systems. An Interactive Multimedia is any computer-delivered electronic system that allows the user to control, combine, and manipulate different types of media (Rockwell and Mactavish, 2004).

2.1.1 VIDEO STREAMING

There are two important parameters that need to be considered in video streaming. One is the Stream and the other is Buffer. Stream is set of bytes and Buffer is matrix of bytes. For example, images are stored in the computer in the form of buffer. Consider an image which is a 4*4 matrix while a Stream will be considered as one row with 16 columns. Suppose we need to download the image, this may take a few seconds and then later image can be presented to the user after it is stored in the buffer. Whereas consider a video data file which has a size of 40 MB or 1GB. If the same technique that is used for images is applied, then the user has to wait for a long period of time and more buffer space will be required to hold the data. To overcome this problem, the technology called streaming is used. Streaming is the technology which allows the user to partly view some part of the video while the other part is downloading (Shridevi Soma *et al*, 2016).

The basic key in video streaming is that the server divides the video into a number of segments and then divided parts are transmitted successively to the client end device. On receiving the video, client stores the received parts of video in its buffer and playback the video in its media player without waiting for the entire large video to be downloaded. It facilitates near instantaneous playback of multimedia content irrespective of the video size (Abinaya and Ramachandran, 2014).

Video streaming can be classified into two categories: live and on-demand. In a live streaming session, a live video content is disseminated to all users in real-time. The video playbacks on all users are synchronized. To the contrary, video-on-demand users enjoy the flexibility of watching whatever video clips whenever they want. The playbacks of the same video clip on different users are not synchronized (Yong Liu *et al,* 2008).

A streaming protocol contains two layers: A transport stream and a control stream. The transport stream carries the media data and it may be based on both transfer protocols, UDP and TCP. The control stream takes care of the Quality of Service (QoS) and its implementation depends on the specific transfer protocol in use. A prominent implementation of a streaming protocol is the Real Time Protocol (RTP). It is based on UDP and uses the Real Time Control Protocol (RTCP) for QoS. The Real Time Messaging Protocol (RTMP) has been published by Adobe in 2009. It uses TCP and guarantees the correct order of the datagrams in the stream. On the server side RTMP implements command messages, data messages, shared object messages to embed e.g. Flash events, media messages, and aggregate messages to combine various messages gaining efficiency. On the client and server side RTMP offers user control messages to notify about events like when the stream is about to begin or end. There are several variants of RTMP that allow to tunnel through firewalls (RTMP-T), to encrypt the media streams (RTMP-E, RTMP-S), and a combination of both (RTMP-TE). Examples for RTMP streaming clients are the VLC player [7] and the browser based FlowPlayer. In response to HTTP requests, the web server sends the Meta file to client. Meta means description. The description of the video stream such as frame rate, size of the video, frame size and so on is contained in the Metafile. So whenever the request for streaming is made by the client, the webserver sends the metafile of the video to establish the connection. This metafile is used by the client's media players for decompression and other tasks. Then a media player sends the RTMP command (Hoecker and Kunze, 2013).

Quality of service (QOS) in a video streaming system is evaluated in three stages: content generation, data delivery and video playback latency (Lou and Hwang, 2012). Content generation refers to the way in which the data is created, data delivery refers to the way the data generated is delivered to the user and video playback latency is the time it takes to reach a downloading peer from an uploading peer during a user's session (Lou and Hwang, 2012).

Recent advances in Peer Assisted Networks, high bandwidth storage devices, compression technology have made providing real time multimedia services over the internet feasible. Real time multimedia has timing setback. For example, video and audio data must be played out simultaneously. If there is a delay in the arrival time of the data, the play out process will pause, which is annoying to human eyes and ears (Dapeng et al, 2001).

2.1.2 VIDEO STREAMING ARCHITECTURE

2.2 NETWORK ARCHITECTURE

Network architecture is the logical and structural layout of the network. It refers to how computers are organized in a system and how tasks are allocated among these computers. Types of network architecture include the client-server network, content delivery network and the peer-to-peer network.

2.2.1 CLIENT-SERVER NETWORK

A client-server network can be defined as a computer network architecture made up of both the client and server, whereby the clients always send requests while the server responds to the requests sent. Client-server provides an inter-process communication because it involves the exchange of data from both the client and server whereby each of them performs different functions. Benefits of the client-server network includes: it splits the processing of application across multiple machines, it allows easier sharing of resources from client to servers, it also reduces data replication by storing data on each server instead of client (Haroon Shakirat Oluwatosin, 2014).

Client-server architecture is usually made up of the application server, database server and computing device. The two main architectures are the 2-tier and 3-tier architecture. The 2-tier client-server architecture involves only the database server and the client computing device, the users will run applications on their device (Client), which connects through a network to the server. The client application runs the coding, and then displays output to the user. It is also called thick client. The 2-tier architecture is considered when the client has access to the database directly without involving any intermediary. The 3-tier architecture involves an application server which serves as a middleware between the client's device and database server. The middleware tier is a separate software running on a separate machine and performs application logic. The client contains presentation logic only, whereby less resources and less coding are needed by the client.

It supports one server being in charge of many clients and provides more resources in the server. Issues with client-server network includes: the servers are quite expensive, the client operating system is easily accessed by servers, and this expose the client system to a number of problems some of which include: physical damage, threats and virus attack (Haroon Shakirat Oluwatosin, 2014).

2.2.2 CONTENT DELIVERY NETWORKS

Content Delivery Networks (CDNs) provide services that improve network performance by maximizing bandwidth, improving accessibility and maintaining correctness through content replication. They offer fast and reliable applications and services by distributing content to cache or edge servers located close to the users. A CDN has some combination of content-delivery, request-routing, distribution and accounting infrastructure. The content-delivery infrastructure consists of a set of edge servers (also called surrogates) that deliver copies of content to end-users. The request-routing infrastructure is responsible for directing client request to appropriate edge servers. It also interacts with the distribution infrastructure to keep an up-to-date view of the content stored in the CDN caches. The distribution infrastructure moves content from the origin server to the CDN edge servers and ensures consistency of content in the caches. The accounting infrastructure maintains logs of client accesses and records the usage of the CDN servers. This information is used for traffic reporting and usage-based billing. In practice, CDNs typically host static content including images, video, media clips, advertisements, and other embedded objects for dynamic Web content. Typical customers of a CDN are media and Internet advertisement companies, data centers, Internet Service Providers (ISPs), online music retailers, mobile operators, consumer electronics manufacturers, and other carrier companies. Each of these customers wants to publish and deliver their content to the end-users on the Internet in a reliable and timely manner. A CDN focuses on building its network infrastructure to provide the following services and functionalities: storage and management of content; distribution of content among surrogates; cache management; delivery of static, dynamic and streaming content; backup and disaster recovery solutions; and monitoring, performance measurement and reporting (Pathan and Buyya, 2010).

The three key components of a CDN architecture are: content provider, CDN provider and end-users. A content provider is one who delegates the URL name space of the Web objects to be distributed. The original server of the content provider holds those objects. A CDN provider is a proprietary organization or company that provides infrastructure facilities to content providers in order to deliver content in a timely and reliable manner. End-users or clients are the entities who access content from the content provider's website. CDN providers use caching and/or replica servers located in different geographical locations to replicate content. CDN cache servers are also called edge servers or surrogates. The surrogates of a CDN are called Web cluster as a whole. CDNs distribute content to the surrogates in such a way that all cache servers share the same content and URL. Client requests are redirected to the nearby surrogate, and a selected surrogate server delivers requested content to the end-users. Thus, transparency for users is achieved. Additionally, surrogates send accounting information for the delivered content to the accounting system of the CDN provider (Pathan and Buyya, 2010).

2.2.2 PEER-TO-PEER NETWORKS

In a Peer-to-Peer (P2P) network all nodes are equal. P2P networks act as a model where each device classified as a peer and simultaneously it may be client or server peer. Thus, a peer may request services from other peers, and provide services to other incoming requests from other peers at the same time on the network. A pure P2P networks consists of only peers whereas other P2P networks depends on centralized servers for finding other peers or depends upon concepts such as special peers termed as super nodes. Super-peer architectures employ the fact that the performance characteristics of the peers i.e. processing power, bandwidth, availability is not evenly spread to all the peers in the network. In a super-peer architecture, a small group of peers most of the responsibilities in the network, e.g., routing task and proxy servers. Some P2P network are overlay networks as the peer themselves get addresses of other participating peers. Types of P2P architecture include centralized P2P architecture, structured P2P architecture and unstructured P2P architecture and the super-peer architecture (Chander Diwakar *et al*, 2011).

The centralized P2P Architecture depends on a centralized entity to place data items within the network. a centralized device gives a directory service to all the peers, therefore forming a star network. All peers in the system have to register their data with the centralized server thereby allowing a convenient way to locate any data or peer in the network by using a physically centralized directory. Instead of the actual data only pointers to decentralized peers are kept at the centralized server therefore lessens the load at the central entity. After finding the applicable data with the help of directory, each peer could communicate directly with other peers that keep the data. In the unstructured P2P Architecture, each peer recursively transmits received request to their neighboring peers (also called neighbors) in an attempt to locate all important points in the network. To avoid missing peers in the network, each peer transmits messages to all other known peers, whether those neighboring peers record the relevant data or not. Structured P2P Architecture: In a centralized architecture, complexity of the linear storage central directory entity is restricted and its suffers from scalability which prevent an unlimited number of employed peers. In an unstructured architecture, the communication overhead caused by flooding message is an important deficit. Thus, an efficient and scalable method requires sub-linear increase in the complexity of storage and retrieval, as more peers added in the network. Structured P2P architectures utilize particular overlay structures to map peers and data items having similar address space, enabling a unique mapping from data items to peers given the current state of the network. To guarantee balanced storage and retrieval loads among the peers, the responsibilities for data items have to be distributed as uniformly as possible (Chander Diwakar *et al*, 2011).

2.2.4 PEER ASSISTED CONTENT DELIVERY NETWORKS

Peer Assisted Content Delivery Networks (PA-CDN) is a hybrid architecture that combines the advantages of both Content Delivery Networks (CDNs) and Peer-to-Peer (P2P) systems (Nasreen Anjum *et al*, 2017). Content can be prerecorded or retrieved from live source, delivery refers to the active retrieval or active transmission of information, the main concern of Content Delivery Networks is the efficient delivery and increased availability of content to the consumer (T. Plagemann *et al,* 2014). The basic design philosophy of peer to peer systems is to encourage users to act as both clients and servers, a peer not only downloads data from the network, but also uploads the downloaded data to other users in the Network. The uploading bandwidth of end users is efficiently utilized to reduce bandwidth burdens otherwise placed on servers (Yong Liu *et al,* 2008). Content Delivery Networks are costly due to the expense of maintaining the servers in different Geo-locations and as the number of peers in the peer-to-peer system increases, security and access becomes a major headache.

The hybrid PA-CDNs can be grouped into two main groups: the centralized architecture and the decentralized architecture. In the Centralized PA-CDNs, new users contact the nearest edge nodes via a request-routing system. The edge servers register the peers and store their meta information such as IP addresses, port of the requested content in a centralized database. In the Decentralized PA-CDNs, the peer-to-peer overlay is managed by the designated Tracker-peers which are selected based on how close they are to the edge server. A peer joining the Decentralized PA-CDNs contacts the tracker-peer to obtain the list of active peers (Nasreen Anjum *et al*, 2017).

Challenges of PA-CDNs are peers availability or peer churn, peers' heterogeneity, locality awareness, startup delay, playback delay, inaccessibility and commercial feasibility challenges (Nasreen Anjum *et al*, 2017). In peer churn, a downloading user session might be affected by a sudden departure of an uploading peer, peers may join or leave the network without any notice. Peers heterogeneity stems from how the network is affected by peers with weak and unstable resources. Peers are heterogeneous in terms of access networks, bandwidth and client end device. Locality awareness is concerned with the different physical network. It has to do with the different network topologies, how the network is connected and the different configurations in the network. Playback and startup delay has to do with lapses that occur while streaming the video, the lapse might occur during the session or at the beginning of the session. Inaccessibility has to do with how peers within a private network can initiate a connection with the peers of a public network but the reverse connection is often complicated by administrative policies. Commercial feasibility challenges focus on the unwillingness of some Internet Service Providers, Content Providers and end-users to participate in Peer Assisted Content Delivery Networks

2.3 SOCIAL MEDIA

2.4 NEUROFUZZY

Neurofuzzy approaches represent an integration of artificial neural networks and fuzzy logic that have capabilities beyond either of these technologies individually (Pantazopoulos *et al*, 1998).

An artificial neural network is a biologically inspired computational model that consists of processing elements (neurons) and connections between them, as well as of training and recall algorithms. The structure of an artificial neuron is defined by inputs, having weights bound to them; an input function, which calculates the aggregated net input signal to a neuron coming from all its inputs; an activation (signal) function, which calculates the activation level of a neuron as a function of its aggregated input signal and (possibly) of its previous state. An output signal equal to the activation value is emitted through the output (the axon) of the neuron. Neural networks have been applied to almost every application area, where a data set is available and a good solution is sought. Neural networks can cope with noisy data, missing data, imprecise or corrupted data, and still produce a good solution. The main characteristics of a neural network are:

- Learning: a network can start with "no knowledge" and can be trained using a given set of data examples, that is, input-output pairs (a supervised training), or only input data (unsupervised training); through learning, the connection weights change in such a way that the network learns to produce desired outputs for known inputs; learning may require repetition.
- Generalization: if a new input vector that differs from the known examples is supplied to the network, it produces the best output according to the examples used.
- Massive potential parallelism: during the processing of data, many neurons "fire" simultaneously.
- Robustness: if some neurons "go wrong," the whole system may still perform well.
- Partial match is what is required in many cases as the already known data do not coincide exactly with the new facts. (Nikola K. Kasabov, 1996).

Neurons connections mimic high learning ability of brain by pattern discovery from examples that is previous inputs. When we build a neural network, it is very important to correctly determine which are the inputs, how many layers we use and what kind of activation function we implement. Thus an artificial neural network is characterized by its architecture, its processing algorithm and its learning algorithm. The architecture specifies the number of neurons and the way they are connected. The processing algorithm specifies how the neural network with a given set of weights computes the outputs for any inputs set. The learning algorithm specifies how the network adapts its weights based on the training inputs sets. The main advantage of artificial neural networks is their adaptability: they are capable to learn from the examples presented to them, often capturing quite subtle relationships between data, that are missed even by trained experts (Florin Aparaschivei *et al*, 2007).

Fuzzy logic is derived from fuzzy set theory, where a fuzzy set is determined by a membership function with a range of values between 0 and 1. Fuzzy Logic is a multi-value logic which allows a better understanding of the result of a statement more approximate than precise in real life. A fuzzy set is built from a reference set called the universal set. The reference set is never fuzzy. Assuming U = $\{x_1,\ x_2, \ldots, x_n,\}$ as the universal set, then a fuzzy set A (A ⊂ U) is defined as a set of ordered pairs: $\{\ (x_i,\ u_A, (x_i)\ )\ \}$ where $x_i \in$ U, u_A : U → [0, 1] is the membership function of A and u_A (x) ∈ [0, 1] is the degree of membership of x in A. (Luis Teran and Andreas Meier, 2010).

One way to represent inexact data and knowledge, closer to humanlike thinking, is to use fuzzy rules instead of exact rules when representing knowledge (Nikola K. Kasabov, 1996). Fuzzy systems are rule-based expert systems based on fuzzy rules and fuzzy inference. Fuzzy rules represent in a straightforward way "commonsense" knowledge and skills, or knowledge that is subjective, ambiguous, vague, or contradictory. This knowledge might have come from many different sources. Commonsense knowledge may have been acquired from long-term experience, from the experience of many people, over many years. A fuzzy system has fuzzy input and output variables, defined by their fuzzy values. Fuzzy rules deal with fuzzy values as, for example, "high," "cold," "very low," etc. Those fuzzy concepts are usually represented by their membership functions. A membership function shows the extent to which a value from a domain (also called universe) is included in a fuzzy concept. Inputs to a fuzzy system can be either exact, crisp values for example, or fuzzy values for example "moderate". Output values from a fuzzy system can be fuzzy, for example, a whole membership function for the inferred fuzzy value; or exact (crisp), for example, a single value is produced on the output. The process of transforming an output membership function into a single value is called defuzzification. Advantage of fuzzy systems is that they are easy to implement, easy to maintain, easy to understand, robust, and cheap. ((Nikola K. Kasabov, 1996).

Fuzzy Inference Systems are employed to mimic human minds for decision making, reasoning and exposition of results in an environment of uncertainty and imprecision. The process of fuzzy inference involves membership functions, fuzzy logic operators, and if-then rules. The Adaptive-Neuro Fuzzy Inference System is a hybrid system that combines the potential benefits of both the methods ANN (Artificial Neural network) and FL (Fuzzy Logic). This system has been employed in numerous modeling and forecasting problems. (Vibha Gaur and Anuja Soni, 2012).

2.4.1 NEUROFUZZY PREDICTION STRATEGY

Prediction (forecasting) is the process of generating information for the possible future development of a process from data about its past and its present development. Three different tasks can be distinguished under the generic prediction problem: Short-term prediction (which is the restricted and default meaning of the word "prediction"), modeling (which is finding global underlying structures, models, and formulas, which can explain the behavior of the process in the long run and can be used for long-term prediction as well as for understanding the past) and characterization (which is aimed at finding fundamental properties of the process under consideration, such as degrees of freedom and so on). The major challenges faced in solving the prediction problem are: establishing whether a process is predictable at all, establishing the type of data and the process subject to prediction, defining the right features for presenting the prediction problem, defining how much past data are required for a good prediction and defining a methodology to test the accuracy of the prediction. Prediction can be carried out based on: past data, Heuristic expert rules (for example, rules which suggest whether to buy or sell depending on the current political situation, economic situation, business growth, exports, and so forth), both past data and heuristic rules. Depending on the information used, the following methods have been applied to solving prediction problems: statistical methods (based on regression analysis and probability analysis Neural networks, which perform pattern matching), rule-based systems (including fuzzy rule-based systems, which represent heuristic expert knowledge; rules might be extracted from past data by using machine-learning techniques), hybrid systems (which make use of past data and expert rules in one system). (Nikola K. Kasabov, 1996).

Mohini p. *et al* stated that Artificial Neural Networks (ANNs) have become very popular, and prediction using ANN is one of the most widely used techniques for rainfall forecasting. This paper provides a detailed survey and comparison of different neural network architectures used by researchers for rainfall forecasting. The paper also discusses the issues while applying different neural networks for yearly/monthly/daily rainfall forecasting. Researchers have used different soft computing techniques like Genetic Algorithm (GA), ANN, and Fuzzy logic for rainfall prediction. The paper suggests that rainfall forecasting can be done using statistical models namely Auto Regressive(AR), Moving Average(MA), Auto Regressive Moving Average(ARMA), Auto Regressive Integrated Moving Average(ARIMA), and Multiple Regression. Rainfall data is non-linear in nature. Amount, frequency, and intensity are three main characteristic of rainfall time series. These values vary from place to place, day to day, month to month and also year to year. Each statistical model has some limitations. AR model regresses against past values of the series. MA model uses past error as an explanatory variable. AR and MA both are suitable for developing models for univariate time series. The AR term only tells the number of linearly correlated lagged observations and is not appropriate for the data having nonlinear relationships. (Mohini p. *et al,* 2015).

Florin Aparaschivei *et al* developed an intelligent system prototype, based on neural networks which will predict net sales value, based on values from precedent months. The system was trained with historical data from monthly balances considered as time series, and it was capable to make predictions about business future evolutions or to present the way business should evolve. Architecturally, a multilayer feed forward neural network which is an interconnection of neurons in which data and calculations flow in a single direction, forward, from the input nodes, through the hidden nodes and to the output nodes was used, with one hidden layer. They chose a hyperbolic tangent function (the relationship between hyperbolic sine and cosine functions or as the ratio of the half-difference and half-sum of two exponential functions in the points) as activation function and backpropagation learning algorithm which is a supervised learning algorithm and adjusting the values of weights in such a manner so as to get the same value as the correct output value of network in training data set is the main objective, with generalized delta rule for speeding up the training. A small size company, which has the main activity domain on informatics services was used as the test sample. (Florin Aparaschivei *et al,* 2007).

2.5 RELATED WORKS

Sawant and Patil designed a social media sharing system which utilizes social media to enable effective bandwidth contribution and scheduling at the users, and employ peer assisted design to distribute video streams with low server cost, stop startup delay and also accomplish high prefetching accuracy. By prefetching the prefix chunks of a video that is likely to be viewed by a user, a short startup delay and smooth playback can be achieved. According to the paper in the online social network, user’s preferences of videos are predicted by: Choosing Videos to Prefetch, they proposed that a user joining the swarm has to choose from a list of videos which videos to prefetch. Social Closeness between the Viewing User and the Source User: Users are more likely to watch videos shared by their friends. Such social preference between friends is evaluated by a straightforward metric, which is the fraction of their common friends over their total friends. Preference of a Source User: According to the paper observations, both historical selection and social closeness predict a viewing user's future preference of source users. Video popularity: The popularity of a video reflects how much viewing users like the video. When a user logs in his account, he is provided a list of unwatched videos shared by his friends and other users. After all this, the user (peer) downloads prefix chunks of the ranked videos in order. When prefetching a video, a peer actively downloads the prefix chunks of that video, i.e., the first several chunks prefetching strategy is carried out by a peer locally when the buffered chunks for the current video can be played. The startup delays thus can be reduced when the user plays videos that have been prefetched.

In 2013, Haiying Shen *et al* stated that Online social networks (OSNs) for example, Facebook, Twitter are now among the most popular sites on the Web and Video sharing has been an increasingly popular application in OSNs, enabling users to share their personal videos or interesting videos they found with their friends. However, OSN’s further advancement is severely hindered by the intrinsic limits of the conventional client/server architecture of its video sharing system, which is not only costly in terms of server storage and bandwidth but also not scalable with the soaring amount of users and video content in OSNs. In an OSN, videos are visited and spread by the users’ friends through the Friend-of-Friend (FOF) relationship. The authors of the paper studied Facebook and their measurement revealed that most of the viewers of a user’s videos are the user’s close friends, most video views are driven by social relationships, and the rest are driven by interests, and viewers of the same video tend to reside in the same location. Based their observations, they proposed SocialTube, a system that explores the social relationship, interest similarity and location to enhance the performance of video sharing in OSNs. SocialTube also incorporates a Social Network based chunk prefetching algorithm to minimize video playback startup delay. Specifically, SocialTube incorporates four algorithms: a Social Network (SN) based P2P overlay construction algorithm, a SN-based chunk prefetching algorithm, chunk delivery and scheduling algorithm, and a buffer management algorithm. (Haiying Shen *et al,* 2013).

It was discovered that as social media becomes increasingly more ubiquitous and influential, the video propagation patterns and users' sharing behavior dynamically change and evolve as well. In this paper, the authors propose a systematic methodology and associated online algorithm for forecasting popularity of videos promoted by social media. Social Forecast algorithm is able to make predictions about the popularity of videos while jointly considering the accuracy and the timeliness of the prediction. We explicitly consider the unique situational conditions that affect the video propagated in social media, and demonstrate how this context information can be incorporated to improve the accuracy of the forecasts. A generic Web 2.0 information sharing system was considered in which videos are shared by users through social media. They assigned each video with an index k ∈ {1, 2, ..., K} according to the absolute time t_{int}^{k} when it is initiated. Once a video is initiated, it will be propagated through the social media for some time duration. They assumed a discrete time model where a period can be minutes, hours, days, or any suitable time duration. A video is said to have an age of n ∈ {1, 2, ...} periods if it has been propagated through the social media for n periods. In each period, the video is further shared and viewed by users depending on the sharing and viewing status of the previous period. It focuses on social media. (Jie xu *e al,* 2014).

2.6 DRAWBACKS

CHAPTER 3

3.1 INTRODUCTION TO THE CHAPTER

In Pshare, users can access their videos based on recommendation systems which have been developed by taking into account user history data and mapping social users' preferences. Pshare uses an Adaptive Neural Based Fuzzy Inference System (ANFIS) to learn users' behavior for video recommendation in online social networks using a Peer Assisted Content Delivery Network (PA-CDN).

3.1.1 SOURCE OF DATA

In Online Social Networks (OSN), videos are visited and spread by the users' friends through the Friend-of-Friend (FOF) relationship. Specifically, a friend of user X, say user Y, receives notification when user X uploads a video, and user Y's friend Z is informed about the video only when user Y ''shares'' the video, and so on. Facebook is one of the most popular sites on the web with over 500 million users. Facebook has options to share videos only between friends, Friend Of Friends (FOFs, default), or everyone. Therefore, users in an OSN watch videos driven much more by the friendship relation than video content. Facebook will be used to get the dataset and the collected dataset will include the information about user friendship relations, interests, locations, and videos uploaded and shared by users and for each video, the video metadata (such as its title, length, and viewers when available).

3.1.2 ADAPTIVE NEURAL BASED FUZZY INFERENCE SYSTEM (ANFIS)

Adaptive Neural-Based Fuzzy Inference Systems or Adaptive neurofuzzy inference systems are fuzzy models put in the framework of adaptive systems to facilitate learning and adaptation. This framework makes fuzzy systems more systematic and less reliant on expert knowledge. Such systems have a network type structure similar to that of an artificial neural networks, which maps inputs through input membership functions and associated parameters, then to outputs. This structure makes these systems very powerful in problems dealing with generalization. Rather than choosing the parameters associated with a given membership function arbitrarily, it would be convenient if these parameters could be chosen automatically so as to tailor the membership functions to the input/output data in order to account for variations in the data values. This is where ANFIS comes into the picture as a solution to this modelling obstacle. To present the ANFIS architecture, let us consider two fuzzy if–then rules based on a first-order Sugeno model:

Rule 1: If x is A_1 and y is B_1, then $f_1 = p_1x + q_1y + r_1$

Rule 2: If x is A_2 and y is B_2, then $f_2 = p_2x + q_2y + r_2$

Here, A_i and B_i are fuzzy sets describing the input and f_i is a crisp (non-fuzzy) variable describing the output. A possible ANFIS architecture to implement these two rules is shown below.

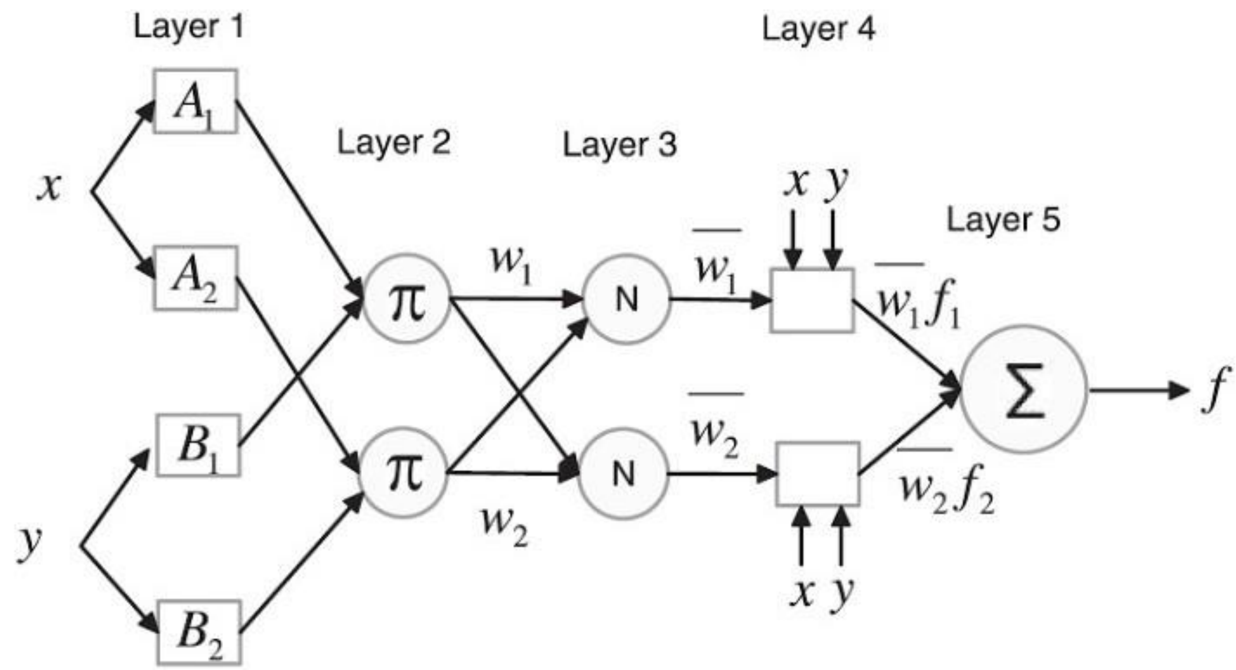

Fig 3.1: ANFIS architecture

Layer 1: O_l,i is the output of the ith node of the layer l. Every node i in this layer is an adaptive node with a node function:

$O_{1,i} = \mu_{Ai}(x)$ for i = 1, 2, or

$O_{1,i} = \mu_{Bi-2}(x)$ for i = 3, 4

x (or y) is the input node i and Ai (or B_{i-2}) is a linguistic label associated with this node. Therefore, the output is a membership grade of the fuzzy set (A_1, A_2, B_1, B_2). The membership function is:

$$\mu_A(x) = \frac{1}{1 + |\frac{x-c_i}{a_i}|^{2b_i}}$$

where a_i, b_i, c_i is the parameter set which are also the precise parameters.

Layer 2: Every node in this layer is a fixed node (not adaptive) labeled π. The output is the product of all the incoming signals. Each node represents the fire strength of the rule.

$O_{2,i} = w_i = \mu_{Ai}(x) \cdot \mu_{Bi}(y)$, i = 1, 2

Layer 3: Nodes in this layer are also fixed nodes. They are labelled N to indicate that they perform a normalization of the firing strength from the previous layer. The output of each node in this layer is given by:

- $O_{3,i} = \overline{w}_i = \frac{w_i}{w_1+w_2}, \quad i = 1, 2$

Layer 4: All the nodes in this layer are adaptive nodes. The output of each node in this layer is simply the product of the normalized firing strength and the parameter set.

- $O_{4,i}$ = wifi = wi(px + qiy + ri)

Where {pi, qi, ri} is the parameter set of this node and are also referred to as consequent parameter.

Layer 5: The single node in this layer is a fixed node labeled ∑, which computes the overall output as the summation of all incoming signals:

- $overall\ output = O_{5,1} = \sum_i \overline{w}_i f_i = \frac{\sum_i w_i f_i}{\sum_i w_i}$

3.1.2 PEER ASSISTED CONTENT DELIVERY NETWORKS (PA-CDNs)

Peer Assisted Content Delivery Networks is a hybrid architecture which combines the Content Delivery Network and Peer to Peer systems. PA-CDNs increase quality of service by relying on contributions of content delivery resources from the end-users and therefore, minimizes infrastructural costs in comparison to traditional CDNs. PA-CDNs reduces the drawbacks of pure P2P systems by replicating the content in a minimal amount of edge servers, therefore, enabling reliable content distribution.

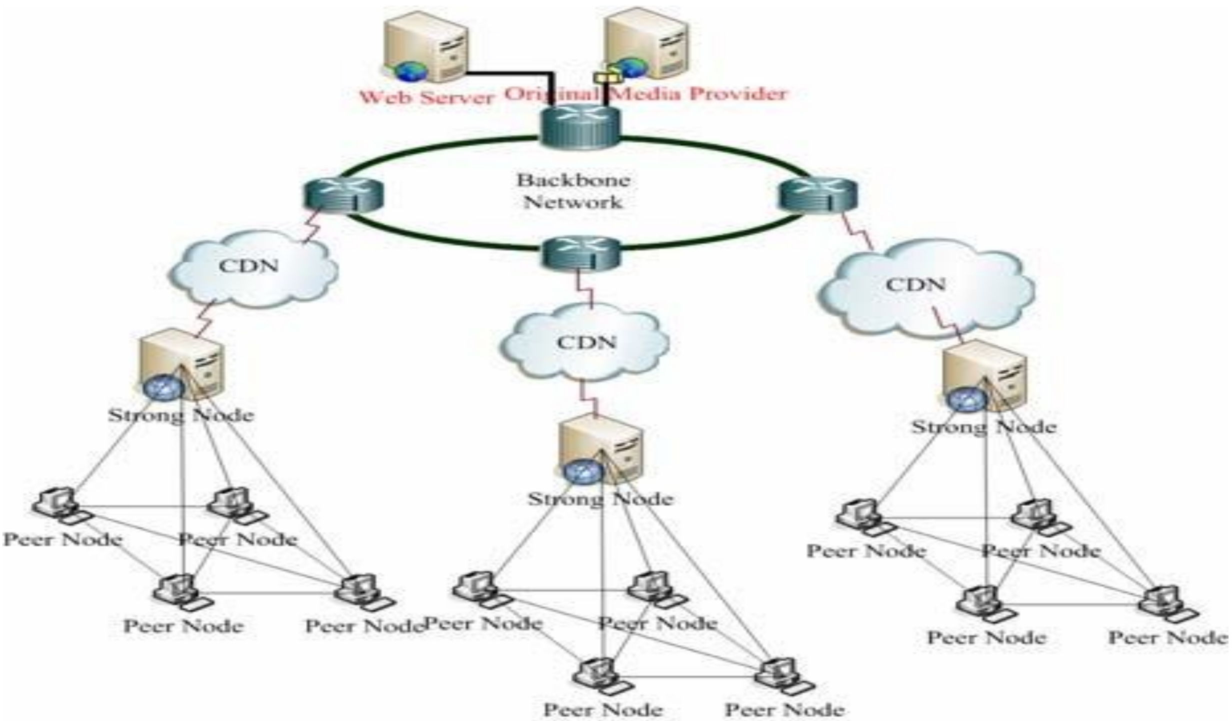

Fig 3.2: PA-CDN architecture

The proposed PA-CDN is a two-layer streaming architecture. The Upper layer is a Content Delivery Network framework layer which consists of the Content delivery servers. The Lower layer is the Peer node layer, which consists of multiple groups of clients who requesting the video streaming services. Each client is considered as a peer node in the group and can act as a potential microserver the other nodes in the group (cluster).

This PA-CDN uses the decentralized architecture in which the peer-to-peer overlay is managed by the designated Strong nodes also known as Tracker-peers which are selected, for instance, based on their proximity to the edge server. A peer joining a decentralized PA-CDN contacts a Tracker-peer to obtain a list of active peers. If the Tracker-Peer is unable to provide a desired list of peers, the request will be redirected to another close-by Tracker-peer until a desired video chunk is found. If sufficient number of peers to download the content is not available, the Tracker-peer will request the missing video chunks from an edge node. Decentralized architectures minimize the required CDN infrastructure by delegating most of the swarm managing functions to the users.

3.2 SYSTEM MODEL

Video sharing in Facebook distinguishes itself from other video sharing websites (e.g., YouTube) in two aspects: video sharing scope and video watching incentives. First, other websites provide system-wide video sharing where a user can watch any video, while in Facebook, videos are usually shared in a 2-hop small circle of friends. Second, users in other video sharing websites are driven to watch videos by interests, while in Facebook, the followers of a source node (i.e, video owner) are driven to watch almost all of the source's videos primarily by social relationship and non-followers watch a certain amount of videos mainly driven by interest.

Pshare establishes a per-node Peer Assisted architecture for each source node. It consists of peers within 2 hops to the source that watch at least a certain percentage of the source's videos. Other peers can still fetch videos from the server. Such peers of a source node S in the social network constitute a cluster for the source node. Based on interest, a hierarchical structure that connects a source node with its socially-close followers, and connects the followers with other non-followers will be built. Thus, the followers can quickly receive chunks from the source node, and also function as a pseudo-source to distribute chunks to other friends. The source pushes the first chunk of its new video to its followers. The chunk is cached in each follower and has high probability of being used since followers watch almost all videos of the source. Further, non-followers sharing the same interest are grouped into an interest cluster for video sharing. We call peers in an interest cluster interest-cluster-peers. A node that has multiple interests is in multiple interest clusters of the source node. Because the source node and followers are involved in every interest cluster for providing video content, we call the group formed by the source, followers, and interest cluster-peers in an interest cluster swarm, and call all nodes in a swarm swarm-peers.

The Artificial Neural Based Fuzzy Inference System will be used to predict interest of non-followers based on history data. The behavior of a user here is called a user profile. The user behavior is learnt from interests and rating values. Rating values in this case represents the videos which the user likes. The user will have multiple interests, each interest will be represented with a number starting from 0, for example a user with 7 interests will have his/her interests represented by 0-6. The interest will make up the first input to the Adaptive Neural Based Inference System. The second input is the user rating. If a user likes a video it will be represented by 1 and if the user does not like the video it will be represented as 0.