ISSN: 0975-833X

Available online at http://www.journalcra.com

International Journal of Current Research
Vol. 7, Issue, 08, pp.19490-19495, August, 2015

INTERNATIONAL JOURNAL OF CURRENT RESEARCH

RESEARCH ARTICLE

Distributed traffic of multimedia files on hadoop

***TaherehShekari, AlirezaDavaryDolatabadi and Abbas Akkasi**

Department of Engineering and Information Technology, Islamic Azad University, E-Campus

ARTICLE INFO

Article History:
Received 07th May, 2015
Received in revised form
07th June, 2015
Accepted 20th July, 2015
Published online 31st August, 2015

Key words:

Multimedia, Audio, Hadoop, Traffic.

ABSTRACT

With spreads usage of practical program, the network traffic increases too. While internet speed increases every day, tendency to benefit from multimedia rise too. Transferring mass size of multimedia data often make a high traffic in network that can be managed by distributed accounting system. In this article optimal solutions for analyzing audio and videotape files are examined.

Citation: TaherehShekari, AlirezaDavaryDolatabadi and Abbas Akkasi, 2015. "Distributed traffic of multimedia files on hadoop", *International Journal of Current Research*, 7, (8), 19490-19495.

INTRODUCTION

Speed of practical programs spreading in multimedia systems of cell phones and computers is high. Most websites broadcast live concerts, online music, TV and radio programs. On the other words, a new class of traffic is created that can transfer a high level of data in every phase. In this new class, not only traffic of network is examined, but also multimedia`s traffic should be consider. This traffic can include sound, music, videotape and image. Rate of data and size of traffic packet play a significant role in identification of multimedia`s general process. In addition, some part of traffic formed by manageable process that all of them, in turn, occupied a percent of bandwidth. Practical programs in client position, decode the digital date and play them by audio player. Usually for coordination of audio coding whit rates of bits, 16 or 20Kb/s is chosen. This process also can increase the size of data. (Bosi and Goldberg, 2012) Most of methods in analyzing network traffic, control by a single server. But if the size of data traffic increase, common procedures` memory, speed of processing and saving capacity will decrease. So we need to manage size of traffic till reduce the pressure of network. In this article we suggest use of Hadoop distributed file system. This file system include high error bearing and low cost in performing on hardware. This system is more practical than last methods.

**Corresponding author: TaherehShekari,*
Department of Engineering and Information Technology, Islamic Azad University, E-Campus

This article will describe traffic distribution of a multimedia file, with the following order. Section 2: significance of multimedia files. Section 3: examination of traffic in network and its analyzing. Section 4: Hadoop distributed file system Section 5: Examination of Hadoop role in decrease of multimedia files traffic.

Study of multimedia files

Multimedia files often have some problems in transmitting information due to their size. For example capturing a film includes mass size of information. Consider a camera that is supposed to record information from space and transfer them to online websites. If we want to send images and movies with that mass size, online sending will be impossible. So, down sampled process performs on it. Data transmitted to processing center of website, then high quality images separated from low quality ones. It means due to mass size of data, we are forced to eliminate some part of data in source, and then in destination restore them by "post process" algorithms. Two of the most important Problems, are bandwidth and traffic. Ability to solve traffic matter, lead to sending high quality data to destination. Most transmitted information to website in form of multimedia or that saved information on memory include a series of long minor links with the same properties. (Samba, 2014) On the other words, a series of similar minor links periodically transfer to website. Transmitting information with low sizes is not so economical. Sending videotapes by connecting nodes

with website depend on size of multimedia files, time of response and delay in transmitting file, website`s power and rate of date loss. (Pauley, 2014)

Scalabiting relies on security of Routing availiity and throughput. Network audio communication is one of the most modern parts of multimedia system under domain of network. Supportive services for audio container, deliver data to a server operational program. Source of this container can be a digital output from audio software for playing an ordinary film or a digital film. Changing channel in TV program or forwarding and repulsing audio file can be some part of these audio online services.

We can call Real audio as one of the online audio systems. This system not only support low size files, but also mass size music. Format of this file (real audio) is RA. Real network described as an open source project. (Humayun *et al*., 2013) Process of Real audio use HTTP. Using HTTP streaming work best with prerecorded file so source alternative protocols have been developed.

The First versions of real audio used 3 different protocols:

1. PNA protocol (progressive network audio) (Liu *et al*., 2014): that is created and used by 5 real systems. PNA are matched with real server`s old versions.
2. RTSP (real time streaming protocol) (Lai *et al*., 2013): they are specifically designed for creating and reading clips. This protocol supports real text and real pix files.
3. HTTP (Hyper text transport protocol): it is secure enough and can cover subject of firewall.

Which used for Meta files that point to real server content. Now we are involve in revolution of multimedia systems, and we seek for a desirable communicative system that in addition to a high quality, be also economical. In general, we can summarize features of online multimedia system as:

a) Using a distributed system instead of narrowband voice telephony for transferring multimedia signals.
b) Changingwireline connection to wireless connection (some things that happen in Viber, Skye, Tango and oovoo)
c) Expanding communications and changing it from people to people communication to people to machine communication.

In clarification of third one, we can refer to visual processing system that it is used in recognition of criminal's faces. In a "people to people" system, images record by camera and transmit directly. However in a smart system, while recording images, act of zooming is done in source, just faces are transmitted and unnecessary parts are eliminated. These cases can be considered in automatic plaque recognition by machine learning algorithms. If we want to transmit a 44k sample /sec and 16-bit audio file, we should transmit 633.3Mb data. But we can use some formats like FLV-MPOG-WMV for decreasing the size of file. (J, M., R, M *et al*., 2012)

Analyzing traffic in network

For increasing processer speed and size of data, more than ever we need to improve communication channel exchanges. One of the most significant matters in network management for better quality is network traffic engineering. Capacity, easy accessibility, communicative cost, high speed and services are significant matters in network exchanges quality.

Exploitation of traffic matrix can define rate of spread use of network. In this matrix, quantity of loaded bar to a node and quality of communication being released. By use of standards in traffic`s models, it is possible to make a model from network traffic by determined input, and then make the traffic a statistic issue model. Here we consider the network structure of IPMPLS. (Fig. 1)

IP traffic at the end of 2015 will reach to a Zeta byte. Universal traffic during last five years became eight-fold as much as before and during next years will grow four-fold.

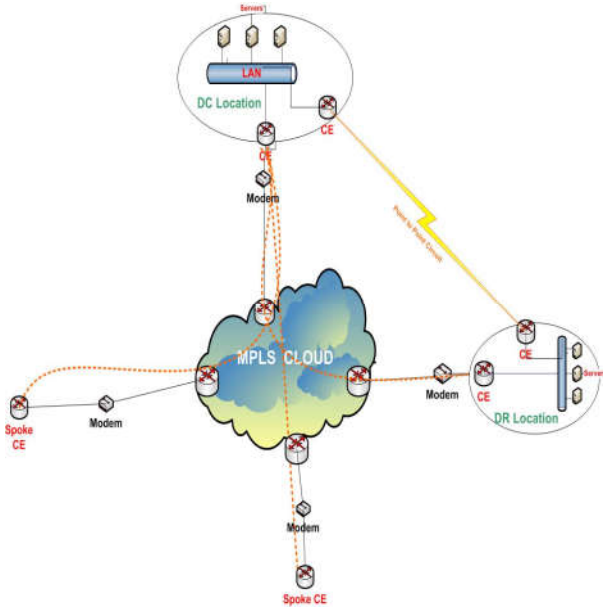

Fig. 1. MPLS Architecture

In general, compound annual growitin of IP traffic is equal to 32%. According to anticipations till end of 2015, every five minutes universal IP network deliver 7.3Peta byte traffic to users. (Motiwala *et al*., 2012) Based on statistics, till end of 2015 traffic of wi-fi system will be 54% and imposed traffic of wire system will reach to 46% (Fig 2).

Furthermore, pick time traffic will get to five-fold, while the mean traffic will get four-fold. And multimedia traffic can categorize in VOIP, Internet Video, etc. A sum of this sectioning is showed in Fig. 3 based on present services.

Furthermore, fixed IP traffic expansion based on TBytes presents in Fig. 4.

Changing process of video, Gawe and their popularity are presented in Fig. 5. Many fold Expansions of traffic in multimedia files is clear in figure.

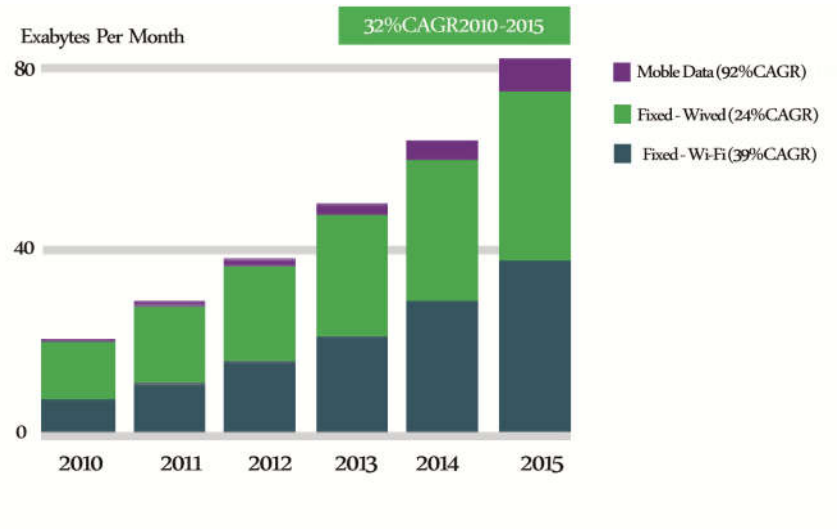

Fig. 2. Comparison of traffic in wireless and wire system 2010- 2015

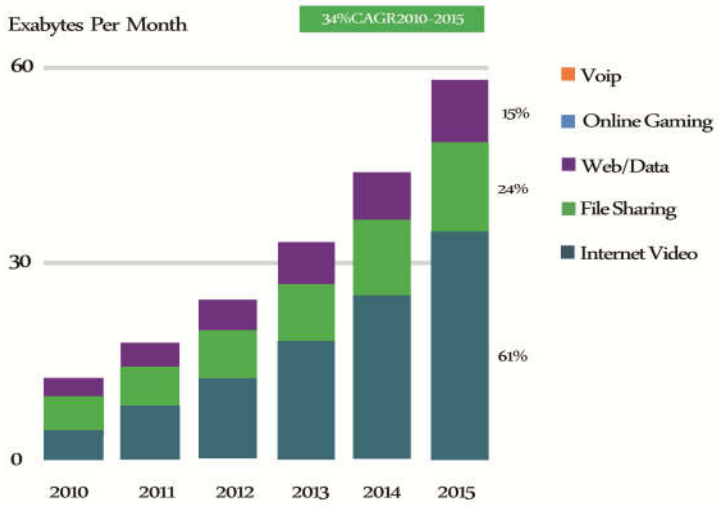

Fig. 3. Comparison of traffic in multimedia system 2010-2015

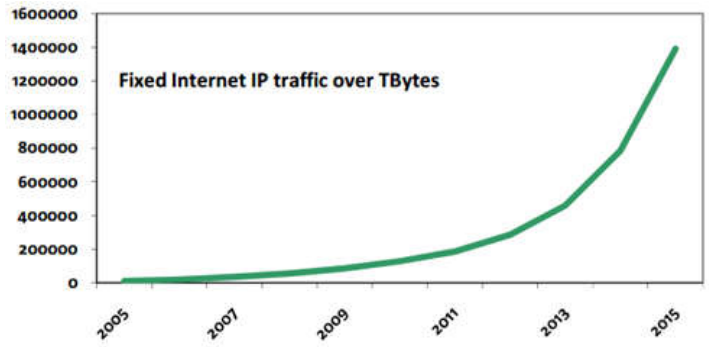

Fig. 4. Expansion of IP traffic over Tera Bytes

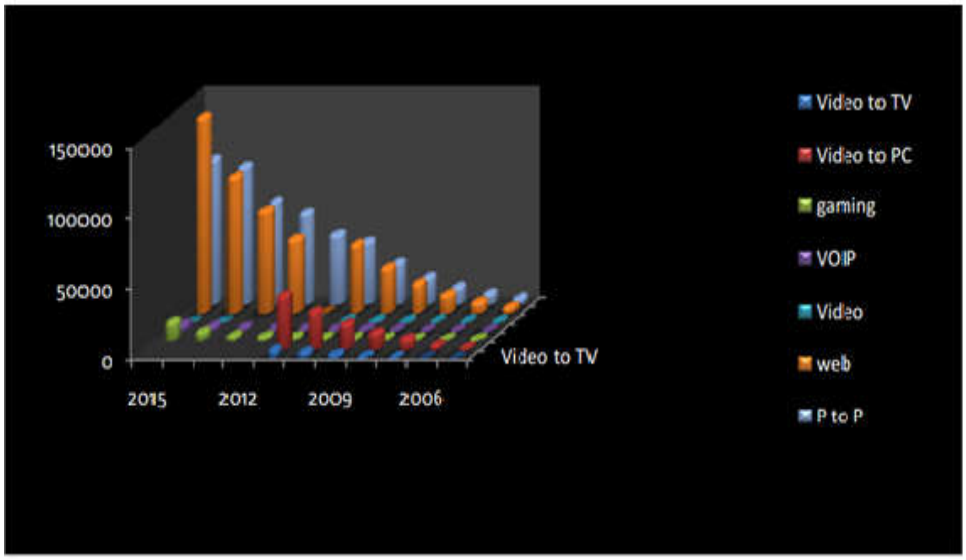

Fig. 5. Changing process of multimedia systems usage

Hadoop distributed file system

Hadoop is a distributed file system with the ability of mass size date saving and processing. This distributed file system provides possibility of common files use that are placed in different physical locations.

Hadoop is based on SQL. With Hadoop, it is possible to save data to the extend of Peta Byte (suitable for multimedia functions). This size is equal with 10billion internet pages. (White, 2012) If we consider every page of internet about 100 kb, usually impairment of any node in distributed file system will disrupt the system, but impairment of node in Hadoop file system can be recognized and repaired automatically by system. Part of Hadoop called Map reduce carry out the organization act and processing files. Well-organizing in Hadoop performs with a high speed. Hadoop is inexpensive because of two reasons:

1. Being open sources
2. Hadoop hardware and architectural processors are slave/master like and every HDFS cluster (Hadoop distributed file system) consists of a Name Node and several Datanodes. HDFS gives the ability of saving user`s data in files. Each file in HDFS is divided to one or many blocks and this blocks save in a set of internal nodes.

Namenode in HDFS is a kind of storage for all metadata. Data don't inter to Namenode. Namenode performs activities like opening, closing and changing name of files. node of data are responsible for reading and writing requests from system file clients. Fig 6 shows HDFS architecture. Namenode and datanode are parts of software that are designed to carry on hardware machines. (Le *et al*., 2013) HDFS supports file hierarchal organization. A user with a program can create folders and save files in it. Hierarchy of system file name space is similar to most of existent system file that have the ability of creating or deleting files or renaming and, or moving a file from one folder to another one. HDFS is more suitable for programs that no need to move constantly.

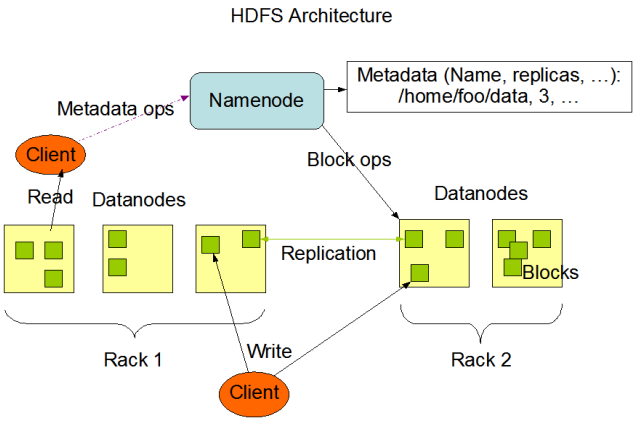

Fig. 6. HDFS architecture

Traffic classification based on Hadoop distributed system

Today, for eruption of practical programs and different services, networks are more complicated and variable than ever. So, network traffic is increasingly developing, but existent methods of network traffic are not expanded for preventing current increasing use of network. Classification of program traffic is the most important part in analyzing network traffic.

Method of classification in programs traffic is a kind of method for identification of data traffic, while comparing with load. Traffic can identify with content and close information. In this article, we present a traffic classification based on mass size multimedia files of Hadoop. Among traffic classification methods, rate of Hadoop based programs is high. But it has some defects in Hadoop based analyzing method, due to the need of high processing bar rather than other methods.

Due to existence of different programs on the internet, classification of programs traffic is one of the most significant issues in network. In the past, programs like SMTP, FTP and HTTP, with below 1024 port number, were dedicated to internet traffic. So traffic classification for getting results based on port information was reliable and accurate, but in today`s network, data port number generates dynamically that can make some problems for classification.

Then, analyzing based on port number cannot guarantee reliability of system like past. For solving this problem, we suggest some methods. Current method can divide in three different kinds: classification method based on signature, classification method based on unity and classification method based on machine learning. As followed:

1. Classification method based on traffic unity, using information for classification of traffic process. This method uses features like system address (IP address, port number, protocol) traffic time of occurrence and type of occurrence.
2. Classification method based on machine learning, that use classification techniques an clustering of machine learning for traffic classification. This method uses features like port number, time duration, arrival time and so on, that can show feature of internet traffic. (Jin *et al*., 2012)
3. In classification based on signature, programs contain a special feature called signature, that separate them from other programs and classification is done by this signature. This method shows accurate results but processing speed and load bar of system are some of weak points of this method.

When a slave node becomes defective, copies prevent to loss data. Several maps manage distributed accounting dependencies and combine reduce dependency of distributed data with several maps. Several maps perform in the same distributed accounting dependencies. A reduce dependency perform for combination of dependencies. So, role of map-reduce in distributed processing of Hadoop system is critical.

Structure of software traffic classification for mass size files in Hadoop system describe as followed. Information system takes the flow as an input and shows size of flow, package and traffic byte as an output. At first, for Hadoop system we perform process of data traffic gathering and input file structure. Then, we present software traffic for classification of Hadoop system. General process of procedure showed in Fig. 7. (Lee and Lee, 2013)

We collect packages of traffic that happen in network. Collecting packages change to FLOW format by producer of the process. Flow defines by five criteria include source IP address, source port, port number, destination IP address and destination port. Input of Hadoop system is a text file. Also, its output is a text file too that uses a FLOW file, based on three extracted conditions. (Luo *et al*., 2014)

1. We only analyze HTTP traffic. Because coding of traffic other than HTTP in analyzing will face some problems.
2. We ignore constant flow. So, we only analyze the first FLOW and withdraw repeated FLOWS. (Luo *et al*., 2014)
3. We only use the first request packet. We cannot analyze all packet, for reduction of analyze accounting size.

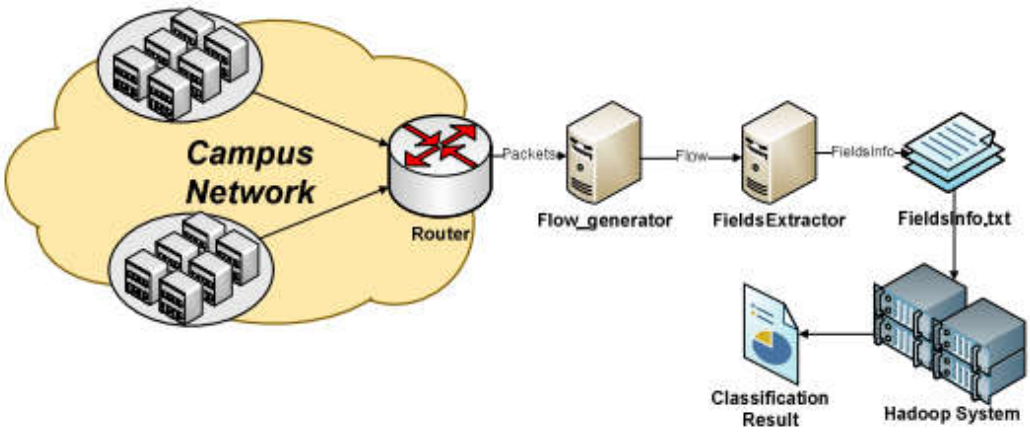

Fig. 7. Application Traffic Analysis System on Hadoop

This article presents a new method in traffic classification of Hadoop system for improving system bar and processing speed in multimedia files. Though there are several researches about reduction of processing speed, but most of them focus on algorithm improvement. Hadoop is a kind of platform that support distributed storing and distributed accounting capacities. Entry files are divided to different segments based on size and distribution among several storing nodes. Date will copied three times, while entering HDFS.

Process of software classification by use of map and reduce dependencies in Hadoop system use FLOW file. Process of software traffic classification in text file and map, reduce dependencies in Fig 8, described below. At first a flow of information center in map dependency. A flow of information consists of eight sections. We are seeking for Payload. Then we compare flow payload and software signature. If they were in agreement , we change software name. Then declare its name as an output map dependency. (Chang *et al*., 2012)

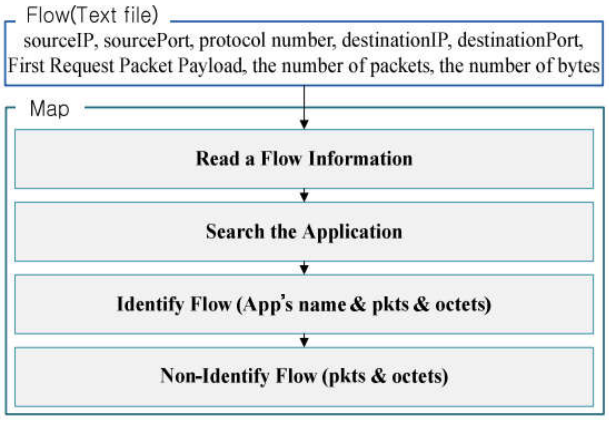

Fig. 8. Process of Map function

Reduce dependency gathers identical transmitted keys. So, may know number of specific software flow. But, there is a distinction in processing, and that is printing output packet number, byte number and number of flow. Hadoop is a kind of platform that support distributed data process. As a increase in network traffic and in data of multimedia system, this system can be a critical one for compatibility of system. A system based on single server in comparison with Hadoop system is more practical in low bar date traffic, so we have some burden between nodes in process of map- reduce for organizing procedures. But speed of processing in reduce-map system is faster. But there are some differences between Hadoop and single server system in experimental environment. Hadoop system analyzes text files that extracted by field extractor module, while single server system analyzes flows with binary frames. So comparison the exact time of processing. When we consider a server based system, while processing duration increases, final extraction time will increase too.

But in a multimedia information interchange system based on Hadoop, while time of data traffic extraction increase, processing duration will not increase too much. So we see that suggested Hadoop system will improve efficiency, while act as an input, even in using binary files.

Conclusion

Today, most websites which present audio and telephonic services, transfer the traffic to server, this technology is a new class of traffic on internet. In this article, we examined increasing traffic under web and found out that traffic follows a series of repetitive models. As soon as user plays a selection audio, flows a mutual information interchange. In addition to data size, manageable information and bandwidth also effect traffic size. On the other hand, we know that over load parts are input and output. In this article, we analyzed Hadoop system because its high error bearing and high speed processing, so by traffic management we process mass size of multimedia data.

REFERENCES

Bosi, M. & Goldberg, R. E. 2012. Introduction to digital audio coding and standards (Vol. 721). *Springer Science & Business Media*.

Humayun, F., Babar, M. I. K., Zafar, M. H., &Zuhairi, M. F. 2013, November. Performance analysis of a Token Bucket Shaper for MPEG4 video and Real Audio signal.In Smart Instrumentation, Measurement and Applications (ICSIMA), 2013 IEEE International Conference on (pp. 1-4).IEEE.

J, M., R, M., K, S. & A, Z. 2012. Encryption of Compressed MultiMedia Data. *International Journal of Computer Applications*, 59(19), 51-55.

Jin, Y., Duffield, N., Erman, J., Haffner, P., Sen, S. & Zhang, Z. L. 2012. A modular machine learning system for flow-level traffic classification in large networks.ACM Transactions on Knowledge Discovery from Data (TKDD), 6(1), 4.

Lai, C. F., Chao, H. C., Lai, Y. X., & Wan, J. 2013. Cloud-assisted real-time transrating for HTTP live streaming. Wireless Communications, IEEE, 20(3), 62-70.

Lee, Y., & Lee, Y. 2013. Toward scalable internet traffic measurement and analysis with hadoop. ACM SIGCOMM Computer Communication Review, 43(1), 5-13.

Liu, W., Mei, T., & Zhang, Y. 2014. Instant mobile video search with layered audio-video indexing and progressive transmission.Multimedia, IEEE Transactions on, 16(8), 2242-2255.

Motiwala, M., Dhamdhere, A., Feamster, N., & Lakhina, A. 2012. Towards a cost model for network traffic. ACM SIGCOMM Computer Communication Review, 42(1), 54-60.

Pauley, J. 2014. U.S. Patent No. 8,832,264. Washington, DC: U.S. Patent and Trademark Office.

Samba, A. S. 2014. U.S. Patent No. 8,868,725. Washington, DC: U.S. Patent and Trademark Office.

White, T. 2012. Hadoop: The definitive guide. "O'Reilly Media, Inc.".

Chang, B. R., Tsai, H. F., Lin, Z. Y., & Chen, C. M. 2012, August. Access-controlled video/voice over IP in Hadoop system with BPNN intelligent adaptation. In Information Security and Intelligence Control (ISIC), 2012 International Conference on (pp. 325-328). IEEE.

Le, H. H., Hikida, S., & Yokota, H. 2013, January. NameNode and DataNode Coupling for a Power-Proportional Hadoop Distributed File System. In Database Systems for Advanced Applications (pp. 99-107).Springer Berlin Heidelberg.

Luo, J., Liang, Y., Gao, W., & Yang, J. 2014, October. Hadoop based Deep Packet Inspection system for traffic analysis of e-business websites. In Data Science and Advanced Analytics (DSAA), 2014 International Conference on (pp. 361-366). IEEE.

****