Global Journals LaTeX JournalKaleidoscope™
Artificial Intelligence formulated this projection for compatibility purposes from the original article published at Global Journals. However, this technology is currently in beta. *Therefore, kindly ignore odd layouts, missed formulae, text, tables, or figures.*

Measurement and Classification of Network Traffic Analysis using Hardoop

Tanmay Paul[1]

[1] Adamas Institute of Technology

Received: 15 December 2016 Accepted: 2 January 2017 Published: 15 January 2017

Abstract

Network traffic can classified as a process which list computer network based on some parameters like port number and protocols into some traffic classes like undesired, sensitivity etc. Traffic can be implemented differently to differentiate the service required for the user for the specific purpose. The large demand of increase in internet users and increase in bandwidth required for various applications are escalating day by day. The traffic data needs to be classified and analyzed with certain tools. Hardoop is the tool which performs the task in very time efficient manner. Hardoop actually run on commodity hardware which process this huge data with hive. Traffic analysis, measurement and classification are done by hardoop based tools at various parameters of packet and flow level. The derived result is used by network administrator for resolving networking related issue. The measurement of internet traffic and analysis has been implemented from long before but the problem is recent years the user in internet has escalated dramatically. We proposed network traffic management system for analyzing internet traffic of multi-terabytes in extensible manner to perform HTTP, ICMP, UDP, TCP and IP.

Index terms— network traffic, hardoop, traffic management and analysis, HDFS, HIVE, IP.

1 Introduction

he collection of different servers, computers, peripherals, devices when connected to one another for secure mean of communication is described as network which is mainly used for sharing data, or as a means of communication. The process of monitoring network traffic involves managing and analyzing network to overcome any discrepancy that might be a problem for the network. The amount of data involve in communication between network is described as network traffic. The network packet [1] mostly comprised of network data which makes the load within the network. The monitoring mainly involves analyses incoming and outgoing packet. The measurement of traffic over a particular network is called traffic measurement. There are basically two types of techniques involved. Firstly the active techniques and the secondly is the passive technique. Active [2]are more accurate and instructive and the main drawback is that it may over crowd the network by infuse with artificial inquest traffic whereas passive [3] run on the background which can be used to implement network analyzing action and the drawback is that supervise on all network [4]. The main challenge of internet traffic measurement is firstly flow statistics computation time and secondly single node failure. To overcome this problem we implement [5] hardoop framework. Hardoop is actually an open source software framework for large set data processing and storage. It provides necessary possibilities of scaling and fault tolerance which are the most important in networking. We also implement Map Reduce model to resolve the inconsistency in between the hardoop data distribution and network monitoring where data is recorded and splinted and dispense them into cluster for individual processing. The related packets may spread across different splits, thus dislocating traffic structures that are essential for network traffic monitoring. In this paper we have proposed a novel method for network traffic measurement and analysis.

2 II.

3 Software Overview

In 1 are efficient of data analyzing but are limited to storage and measurement. The traffic sampling method can be used to overcome the limitation where results are drawn through partial observation. The implementation of SQL is also not proposed due to its nature of query operation. Below in Table 1 networking traffic monitoring tools are given with uses and limitations are described.

4 System Overview

The system proposed involved firstly input conversion, secondly hardoop pre-processing and qlikView [7] analysis. At first for the packet capture jpcap and wincap [8][9] is used for capturing which is used for supporting the jdk environment and wincap supports the window environment. After capturing the packet gets converted into .text file or .csv file for training data. The dataset made gets loaded as input for category. The processed file is stored in HDFS and to represent in HIVE file externally. And at last IP analysis, port no, protocol and displayed in graphical format. Below in Fig2 the work flow diagram of the proposed system has been given.

5 Experimental Evaluation

Protocol based network packet are captured, port number having LAN making use of java API.2 and IP addresses. The captured file stored in HDFS [10][11] is described data wise. The top 10 IP address can be calculated to define the user usage so that the network which consumes more traffic or more bandwidth can be identified. The total number of packet has also been calculated based on port which his called port-wise byte counts. Port 443 (HTTPS) having the highest number of count which is about 59% has also been shown below. The size of packet and total number of packet each day has been calculated. Below in Fig4 the top 10 IP address usage is shown and in Fig5 the port wise byte count is also shown. V.

6 Conclusion

The network traffic analysis we proposed in this paper will be very efficient for the network administrator to monitor the bandwidth consumption and maintain the system and trouble shoot bugs if necessary. In the paper our main focus was on the flow packet and analysis by network topology. The huge amount of data cannot be handled with single server so large dataset is necessary for matching the computing and storage, and scalable analysis becomes a problem. That the reason we introduce Hardoop as an open-source platform which resolves all the issue in large data set analysis. We have proposed the novel method of data analysis and measurement.

[1] [2]

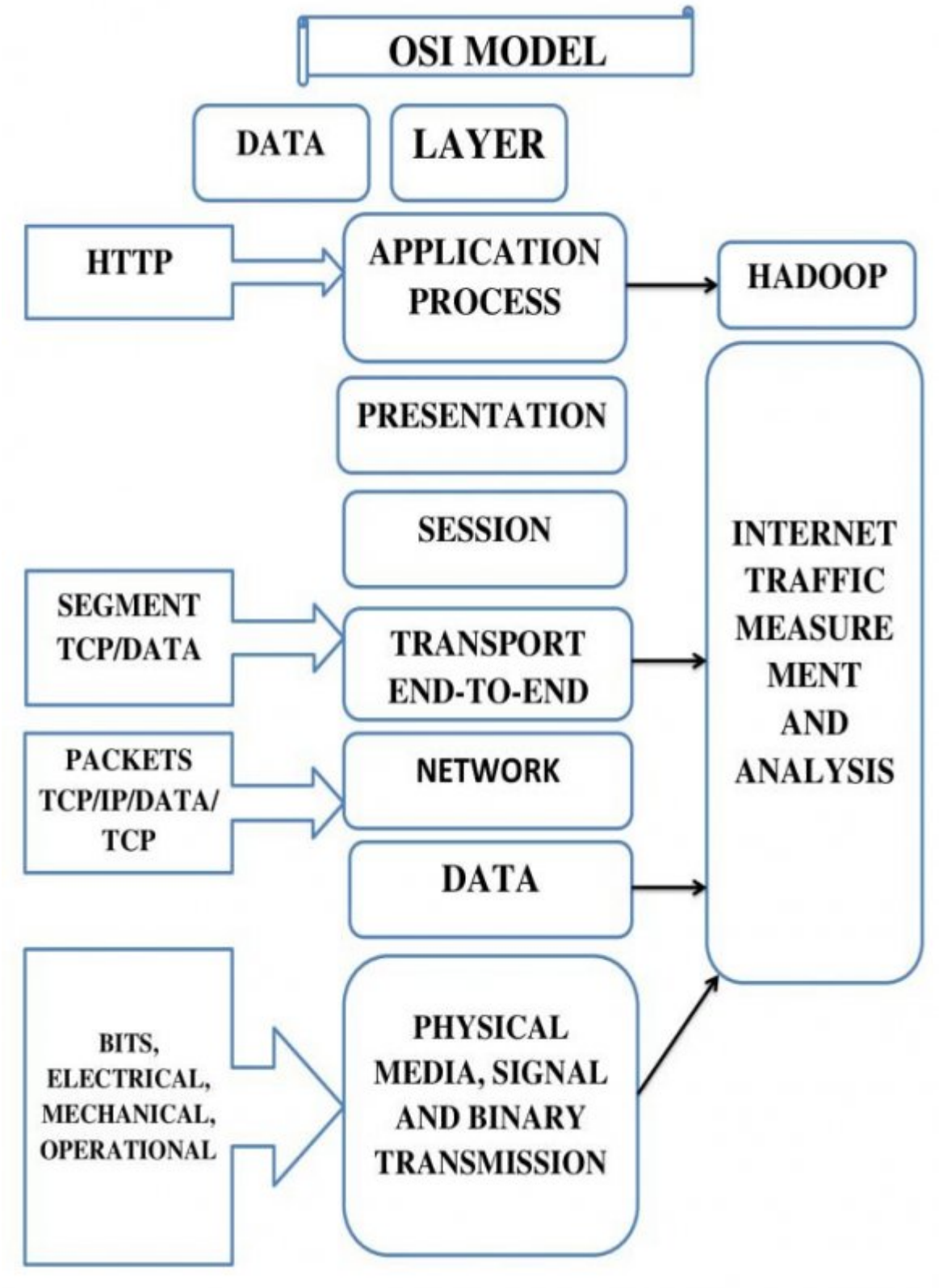
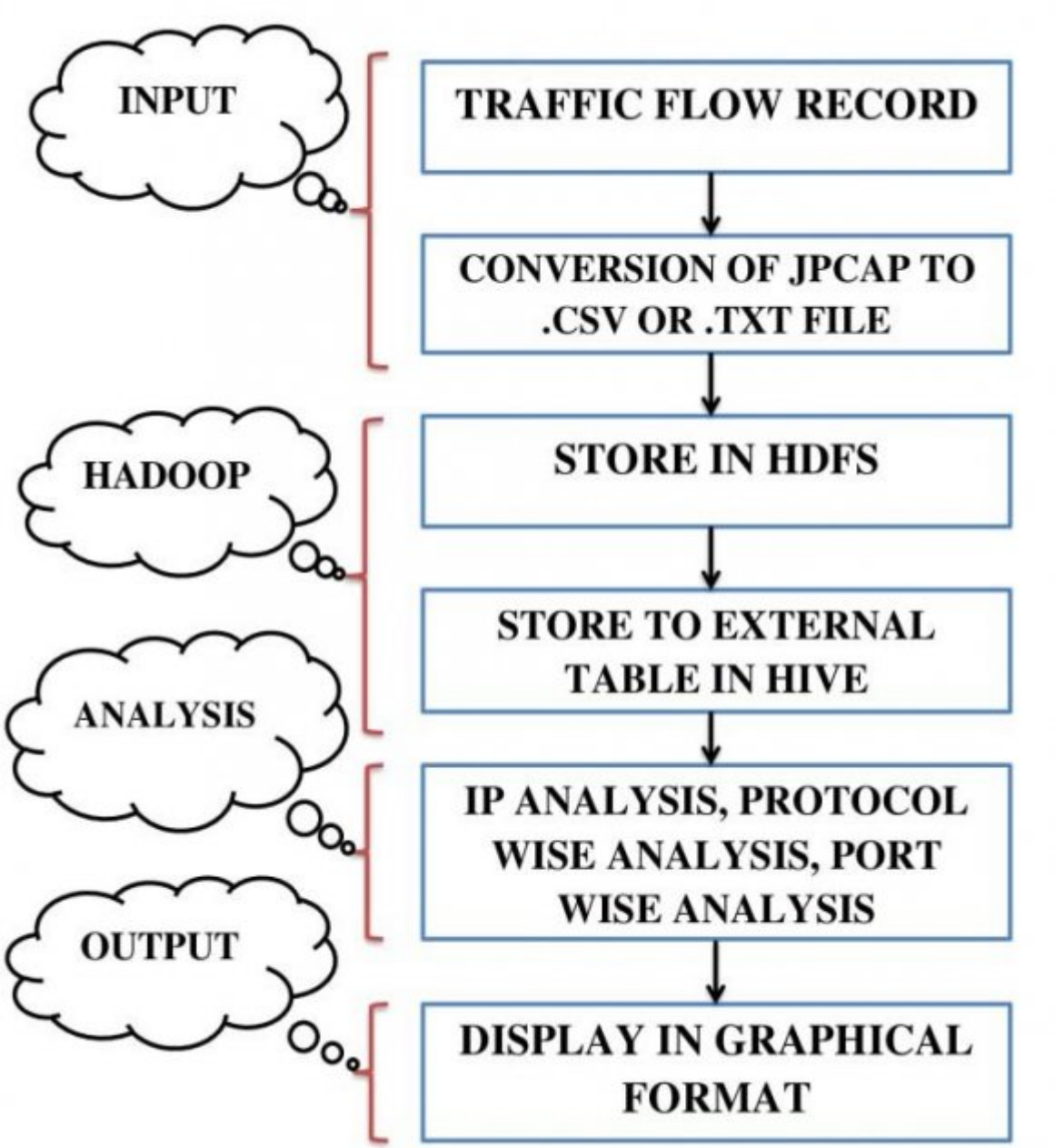

OSI MODEL
DATA
LAYER
HTTP
APPLICATION PROCESS
HADOOP
PRESENTATION
SESSION
SEGMENT TCP/DATA
TRANSPORT END-TO-END
PACKETS TCP/IP/DATA/ TCP
NETWORK
DATA
BITS, ELECTRICAL, MECHANICAL, OPERATIONAL
PHYSICAL MEDIA, SIGNAL AND BINARY TRANSMISSION
INTERNET TRAFFIC MEASUREMENT AND ANALYSIS

Figure 1:

TOOL NAME	OPERATING SYSTEM	LANGUAGE	USE	DISADVANTAGE
NETWORK MINER	Windows, Mac, Linux, FreeBSD.	C	Used as passive network sniffer/packet capturing tool in order to detect OS sessions, host name, open ports etc.	Cost is high about 70$
WIRESHARK	Linux, OS X, BSD, Solaris, windows.	C, C++	It allows examining of the data from a live network or from a capture file on disk. The data can also be browsed and delving down into packet level as required	The main issue is the security features of this tool.
TEPDUMP	Unix like OS, Linux, OS X, BSD, Solaris, windows, android and AIX	C	The user with the necessary privilege acting on a router or gateway through which unencrypted traffic such as telnet passes can use Tepdump to view login id, passwords, URLs, content of website being viewed or any other unencrypted information	It does not receives new features update and keep resolving the bugs and troubleshooting the previous networking issues.

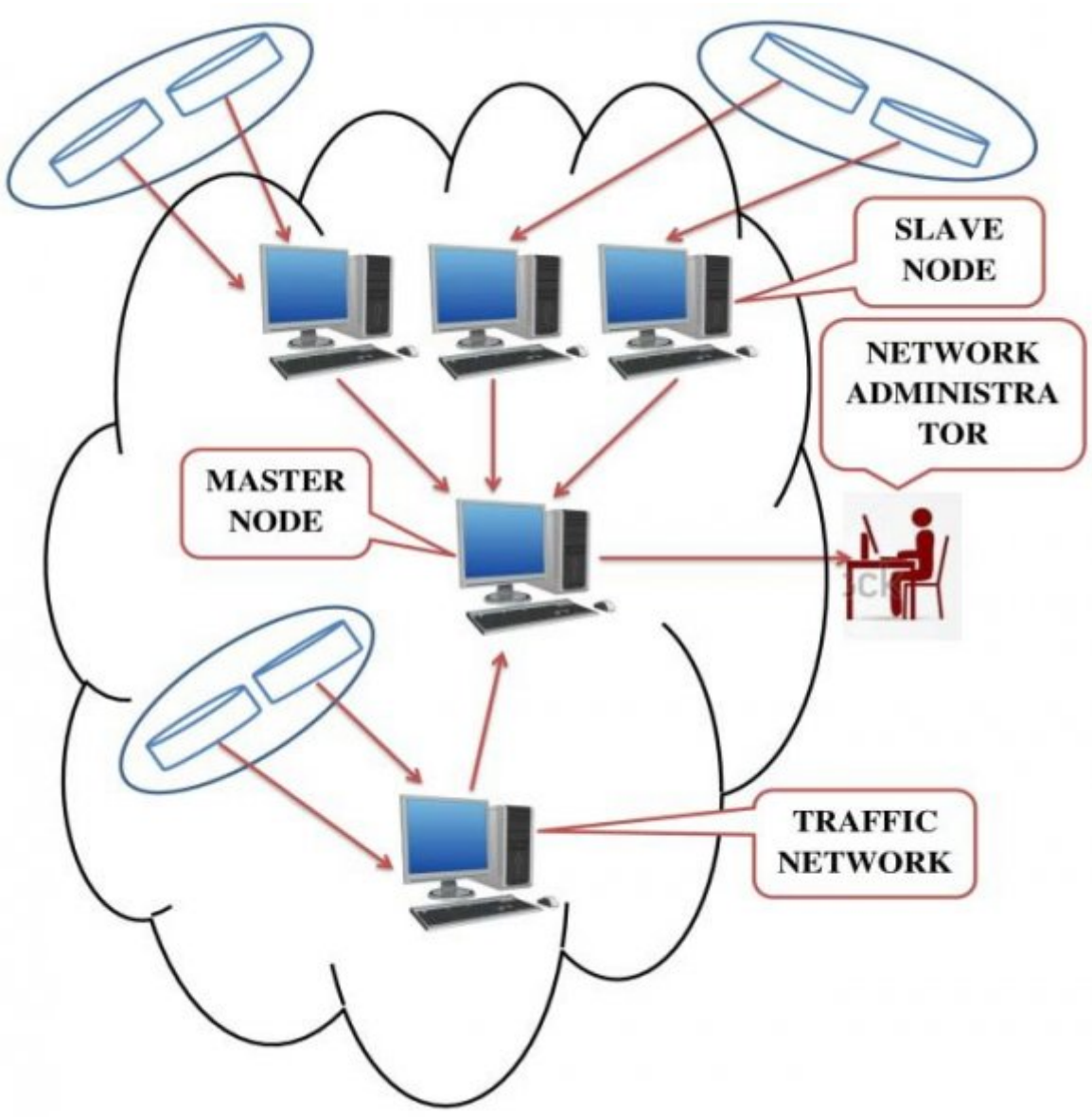
91

Figure 2: T 9 Fig. 1 :



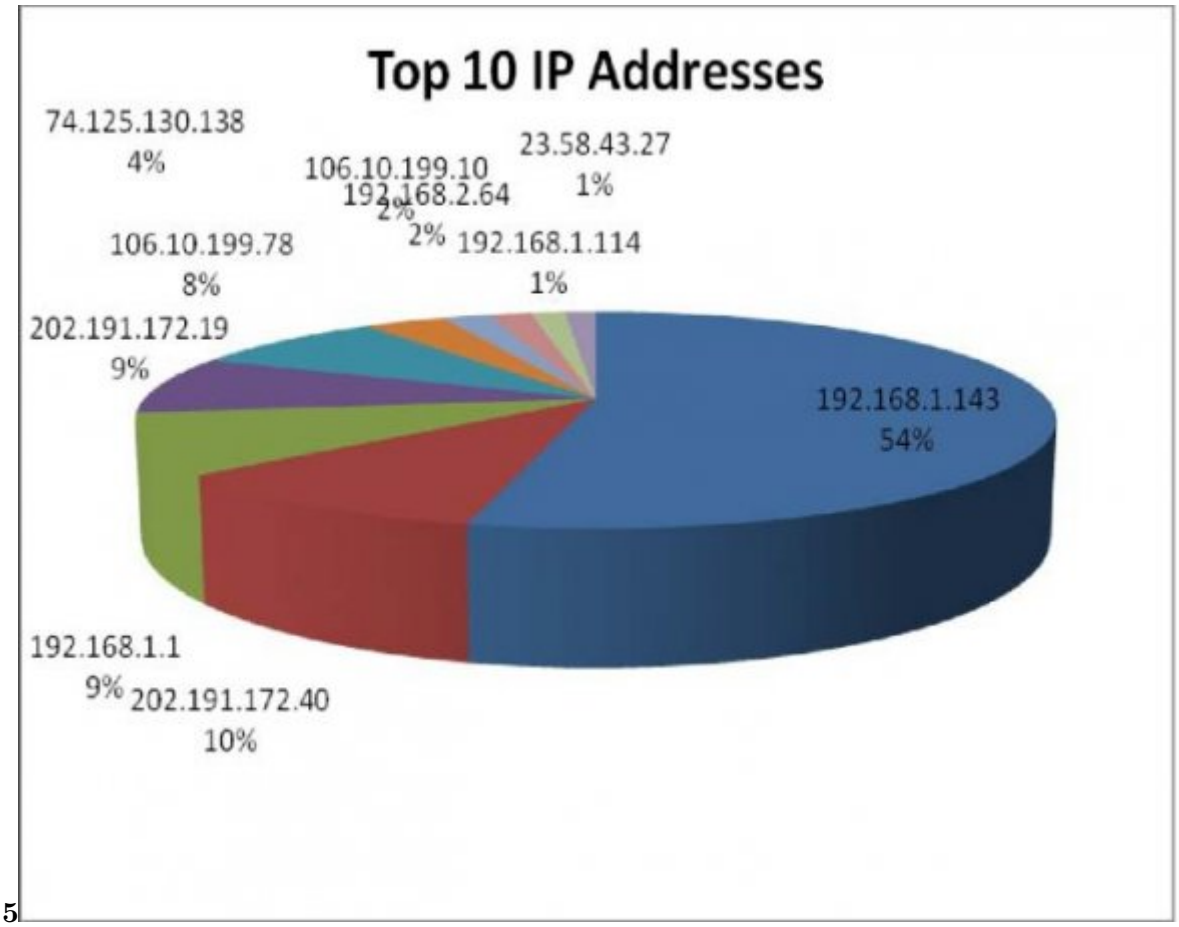
23

Figure 3: Fig. 2 :Fig. 3 :



4

Figure 4: Fig. 4 :



5

Figure 5: Fig. 5 :

1

III.

Figure 6: Table 1 :

[Wu Kehe] , Wu Kehe .

[Rui] , Cheng Rui .

[Ahmadi; Laura Kane] , Ali R Ahmadi; Laura Kane .

[Macdonald] , Robert Macdonald .

[Grahamault and Almeida] , Pedro Grahamault , Almeida . (Sotiris Georgiopoulos)

[Shang et al.] , Shuonan Shang , ; Yongqingmeng , ; Jian Wang .

[Fanibhare] , Vaibhav Fanibhare .

[Roumianailieva and Kirilanguelov] , ; Roumianailieva , Kirilanguelov .

[Yong Xing] , Yong Xing , Wang .

[Ye] , Mao Ye .

[Ibrahim; Rosilah Hassan; Kamsuriah Ahmad and Asrulnizamasat] 'A study on improvement of internet traffic measurement and analysis usingH adoop system'. Lena T Ibrahim; Rosilah Hassan; Kamsuriah Ahmad , ; Asrulnizamasat . *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*,

[Wang and Zhang] 'Accelerating I/O Performance of SVM on HDFS'. Jun Wang , ; Jiangling Yin; Xuhong Zhang . *2016 IEEE International Conference on Cluster Computing (CLUSTER)*,

[Papadopoulos ()] *Active network management supporting energy storage integration into system, market and the distribution network*, Josebarros; Panagiotis Papadopoulos . 2016. CIRED Workshop.

[Mingbo et al.] 'Design and implementation of IP network traffic monitoring system'. Liu Mingbo , ; Sun Wenjie; Zhao Qianhong; Tian , Zhaoping . *2016 15 th International Conference on Optical Communications and Networks (ICOCN)*,

[Wang] 'Design and simulation on the PC of routing software based on Wincap'. Xiu Zhu Jiang; Chun Wang . *The 2011 IEEE/ICME International Conference on Complex Medical Engineering*,

[Gashurova ()] 'Monitoring and optimization of e-Services in IT Service Desk Systems'. Delyana Gashurova . *19th International Symposium on Electrical Apparatus and Technologies (SIELA)*, 2016.

[Li; Wei Ren; Xifan and Cui ()] *Research on modeling and control strategy of modular multilevel matrix converter supplying passive networks*, Huixuan Li; Wei Ren; Xifan , Wang; Yong Cui . 2016. IEEE PES Asia-Pacific Power and Energy Engineering Conference (APPEEC)

[Agrawal; Raymond Hansen; Chunming Rong; Tomasz and Wiktorski] 'SD-HDFS: Secure Deletion in Hadoop Distributed File System'. Bikash Agrawal; Raymond Hansen; Chunming Rong; Tomasz , Wiktorski . *2016 IEEE International Congress on Big Data*, Big Data Congress.

[Dahake ()] 'Smart Grids Map Reduce framework using Hadoop'. Vijay Dahake . *3rd International Conference on Signal Processing and Integrated Networks (SPIN)*, 2016.

[Zhang Yingqiang and Hongtao ()] 'The research on the software architecture of network packet processing based on the manycore processors'. ; Mu Zhang Yingqiang , Hongtao . *7th IEEE International Conference on Software Engineering and Service Science (ICSESS)*, 2016.

[Xing and Li] Wenjian Xing , ; Yunlan Zhao; Tonglei Li . *2010 Second International Workshop on Education Technology and Computer Science*,