

Study and Analysis of Big Data Security Analytics for Protecting Cloud Based Virtualized Infrastructures

Hilal Ahmad Khan¹, Gurinder Pal²

¹M.Tech Scholar, ²Assistant Professor

^{1,2}Department of Computer Science Engineering, IET, Bhattal, Technical Campus, Ropar, Punjab, India

ABSTRACT

In cloud computing virtualized infrastructures has become a stimulating target for cyber attackers to initiate advance attacks. The motive of this work may be a narrative huge knowledge primarily based security analytics approach to get advanced attacks in virtualized infrastructures. User application logs and network logs collected consistently from the tenant virtual machines (VMs) are saved within the Hadoop Distributed File system (HDFS). Extraction of attack features is performed through graph-based event correlation and Map Reduce parser based identification of potential attack paths. Two-step machine learning approaches logistic regression and belief propagation are used to perform the determination of attack presence.

KEYWORDS: HDFS, VM, SIEM, IDS

How to cite this paper: Hilal Ahmad Khan | Gurinder Pal "Study and Analysis of Big Data Security Analytics for Protecting Cloud Based Virtualized Infrastructures" Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-4 | Issue-1, December 2019, pp.763-767, URL: www.ijtsrd.com/papers/ijtsrd29709.pdf



Copyright © 2019 by author(s) and International Journal of Trend in Scientific Research and Development Journal. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0) (<http://creativecommons.org/licenses/by/4.0>)



I. INTRODUCTION

Virtualized framework has become a noteworthy target for cyber attackers to initiate advance attacks. The motive of this work may be narrative massive information based mostly security analytics approach to find advanced attacks in virtualized infrastructures. This approach relies on HDFS (Hadoop Distributed File System) that saves Network and application records consistently from the resident virtual machines. A virtualized framework contains virtual machines (VMs) that rely upon the software defined multi-instance theme of the current hardware. The virtual machine monitor conjointly referred to as hypervisor, assist, and monitor and manages the software defined multi-instance theme. The capability to pool completely different computing measures further as permit on-demand resource scaling has semiconductor diode to the overall distribution of virtualized framework as a crucial equipment to cloud computing services. Security analytics applies analytics on the {various} records that are acquire at various points within the network to find attack existence. By grasping the huge variety of records created by totally different security systems (e.g., intrusion detection systems (IDS), security data and event management (SIEM), etc.), pertain huge information analytics are going to be apt to note attacks that aren't set via signature or rule-based recognition ways. whereas security analytics eliminate necessitate for signature info by applying event association to find already undiscovered attacks, this is often typically not performed in

period of time and current execution are just about not adjustable. This digitalization of the marketing world is swing corporations at hazards of cyber-attacks on top of ever antecedent. Huge information analysis has the potential to supply security verses these attacks. Since the construct of a company security circumference has nearly depart in recent years due to the increasing promotion of cloud and mobile services, data security has tailored a smart pattern switch from regular perimeter security tools relating to detective work and observation distractive activities within company networks. Growing advanced attack technique employed by cyber criminals and therefore the growing task of distractive insiders in varied current large-scale security rupture clearly specify that ancient perspective to data security will now not sustain. Analytics is that the essential part in resistance cyber sturdiness. With growing advanced associated constant attacks and therefore the straightforward incontrovertible fact that each administration should secure itself against all classes of attacks whereas an assailant solely wants one victorious try, company should revise their cyber security plan. They need to plan on the far side pure interruptions towards the PDR pattern: stop – find – Respond. At the key of this proposal stands improved observation which is wherever huge information analytics comes into play. Recognition should be able to distinguish dynamic use influence to implement compound review quickly, near real time; to execute

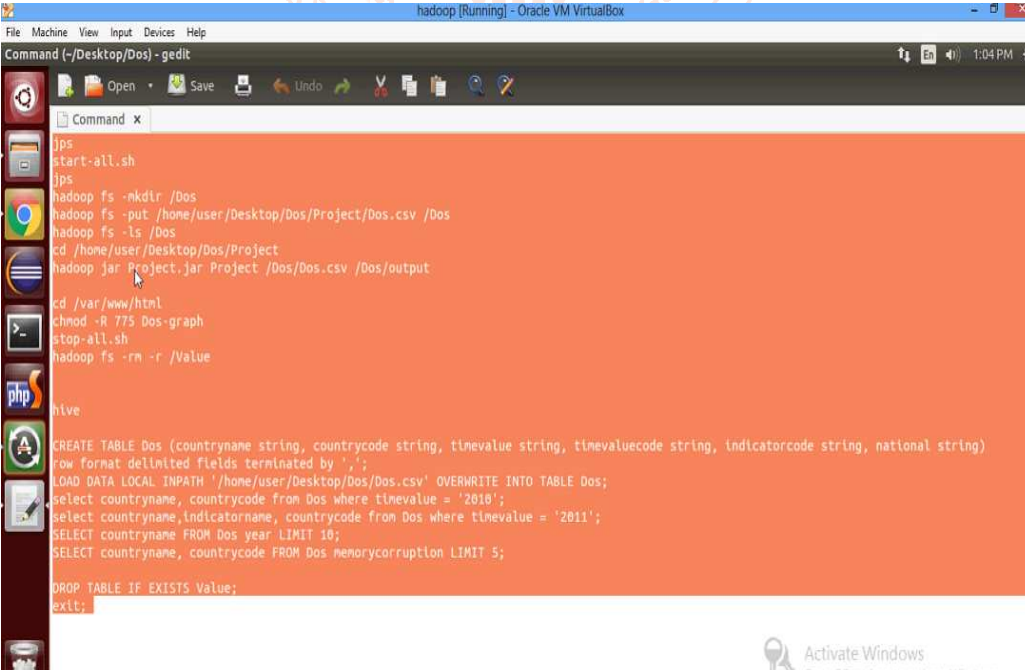
complicated correlations over a range of information origin vary from application logs and server to network events and user activities. This would like each leading analytics on the far side straightforward rule-based perspective and therefore the capability to run analysis on huge quota of gift and historical information – big data security analytics. Mingle the current condition of analytics with security helps administration upgrade their cyber flexibility. Because the security industry's feedback to those disputes, a brand new origination of security analytics solutions has appeared in current years, that are able to gather, save and examine monumental amounts of security information over the full venture in real time. Upgraded by further subject information and external warning intelligence, this information is then examining victimization totally different correlation algorithms to note deviation and thus acknowledge doable vicious activities. In contrast to classic SIEM solutions, aforementioned tools utilize in close to real time and make a tiny low quantity of security attentive align by intensity consistent with a threat model. These alerts are improved with further argumentative details and are able to deeply clarify a security analyst's job and authorize fast awareness and reduction of cyber-attacks.

II. PROBLEM DEFINITION

Virtualized infrastructure in cloud computing has become an attractive target for cyber attackers to launch advanced attacks. The existing techniques for protecting cloud based virtualized infrastructure include malware detection and security analytics. Malware detection usually involves two steps, first, monitoring hooks are placed at different points within the virtualized infrastructure then a regularly-updated attack signature database is used to determine attack presence. While this allows for a real-time detection of attacks, the use of a dedicated signature database makes it vulnerable to zero-day attacks for which it has no attack

IV. RESULTS

The results obtained are as under:



```

File Machine View Input Devices Help
hadoop [Running] - Oracle VM VirtualBox
Command (-/Desktop/Dos) - gedit
Command x
jps
start-all.sh
jps
hadoop fs -mkdir /Dos
hadoop fs -put /home/user/Desktop/Dos/Project/Dos.csv /Dos
hadoop fs -ls /Dos
cd /home/user/Desktop/Dos/Project
hadoop jar Project.jar Project /Dos/Dos.csv /Dos/output
cd /var/www/html
chmod -R 775 Dos-graph
stop-all.sh
hadoop fs -rm -r /Value
hive
CREATE TABLE Dos (countryname string, countrycode string, timevalue string, timevaluecode string, indicatorcode string, national string)
row format delimited fields terminated by ',';
LOAD DATA LOCAL INPATH '/home/user/Desktop/Dos/Dos.csv' OVERWRITE INTO TABLE Dos;
select countryname, countrycode from Dos where timevalue = '2010';
select countryname, indicatorname, countrycode from Dos where timevalue = '2011';
SELECT countryname FROM Dos year LIMIT 10;
SELECT countryname, countrycode FROM Dos memorycorruption LIMIT 5;
DROP TABLE IF EXISTS Value;
exit;
  
```

Fig.1: Different commands used in work

signatures. Security analytics applies analytics on the various logs which are obtained at different points within the network to determine attack presence. By leveraging the huge amounts of logs generated by various security systems like (intrusion detection systems (IDS), security information and event management (SIEM), applying big data analytics will be able to detect attacks which are not discovered through signature or rule-based detection methods. While security analytics removes the need for signature database by using event correlation to detect previously undiscovered attacks, this is often not carried out in real-time and current implementations are intrinsically non-scalable. Hence need a novel big data security analytics approach to protect virtualized infrastructures against advanced attacks.

OBJECTIVES

- To study and analyze different security techniques.
- To propose a novel Big Data Security Analytics approach for protecting virtualized cloud infrastructures from advanced cyber-attacks.
- To make use of HDFS (Hadoop Distributed File System) for implementing high end security.

III. METHODOLOGY

The proposed work includes a novel big data based security analytics (BDSA) approach to protecting virtualized infrastructures against advanced attacks. With the usage of network logs and user application logs obtained from the guest virtual machines stored in HDFS (Hadoop Distributed File System), our proposed BDSA approach first extracts attack features through graph-based event correlation, a Map Reduce parser based identification of potential attack paths and then ascertains attack presence through two-step machine learning, namely logistic regression and belief propagation.

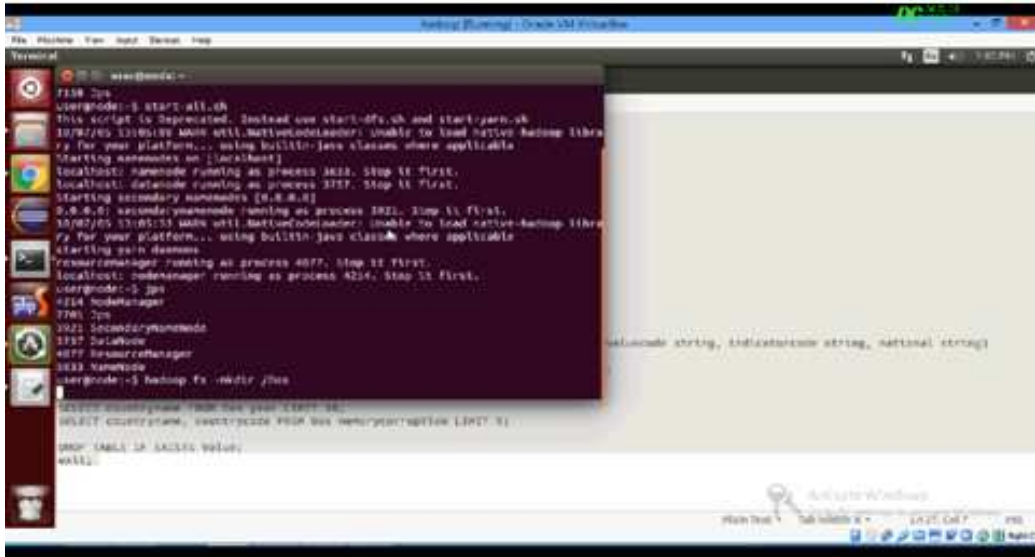


Fig.2: Hadoop make directory

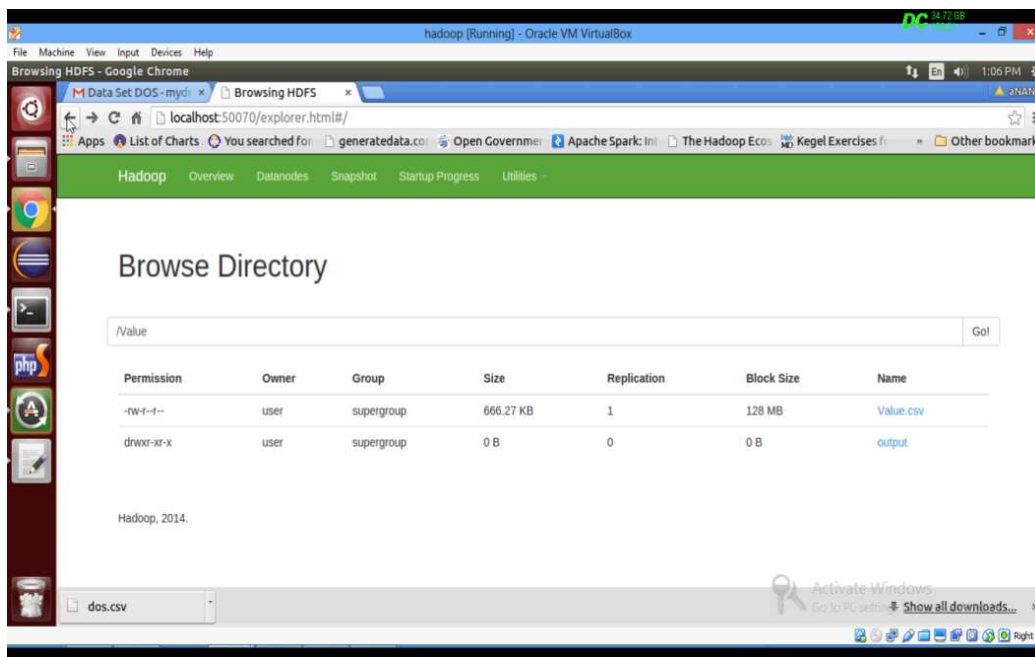


Fig.3: Hadoop file system

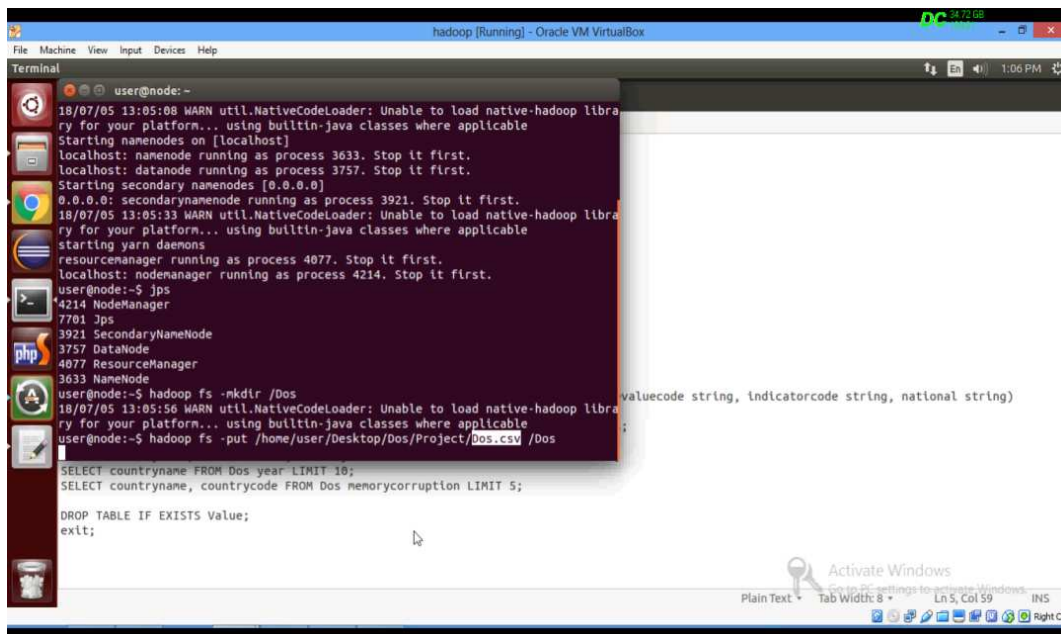


Fig.4: Hadoop commands

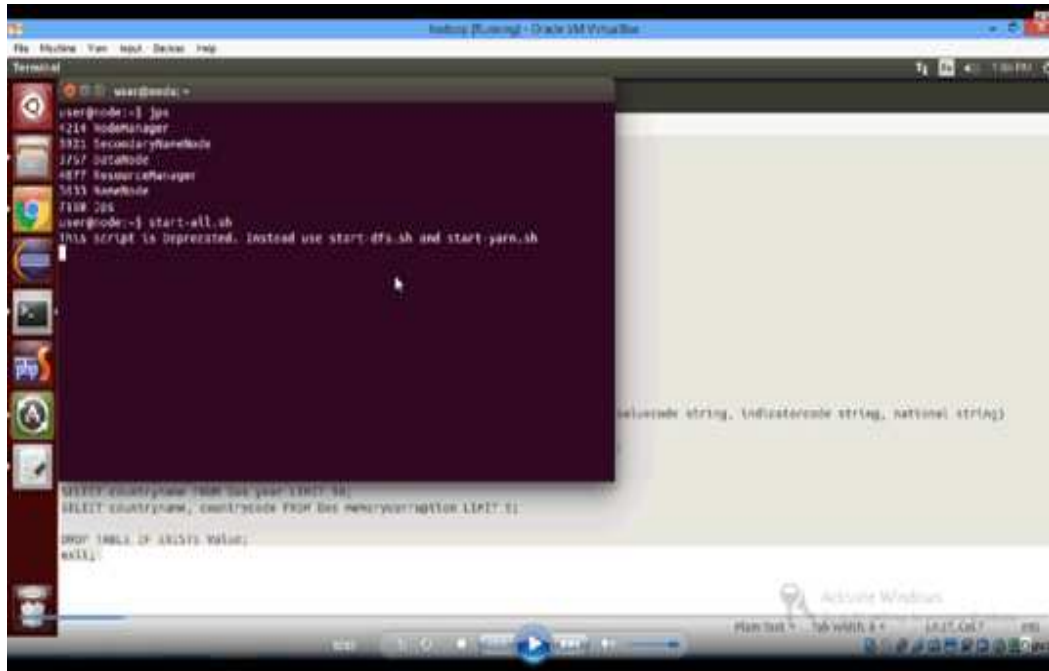


Fig.5: Main menu of demo work

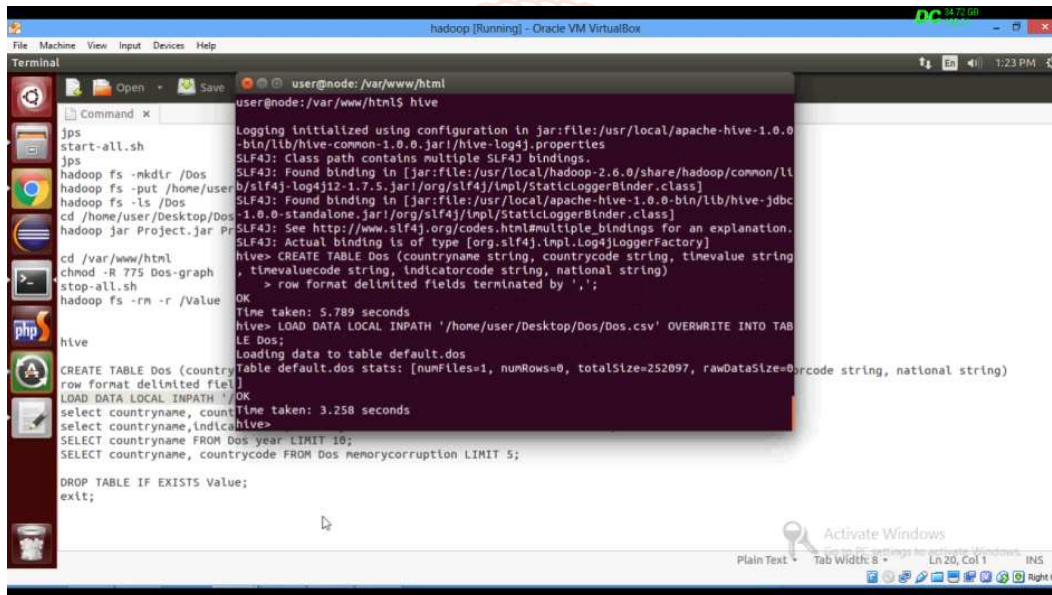


Fig.6: Loading test file with Hive

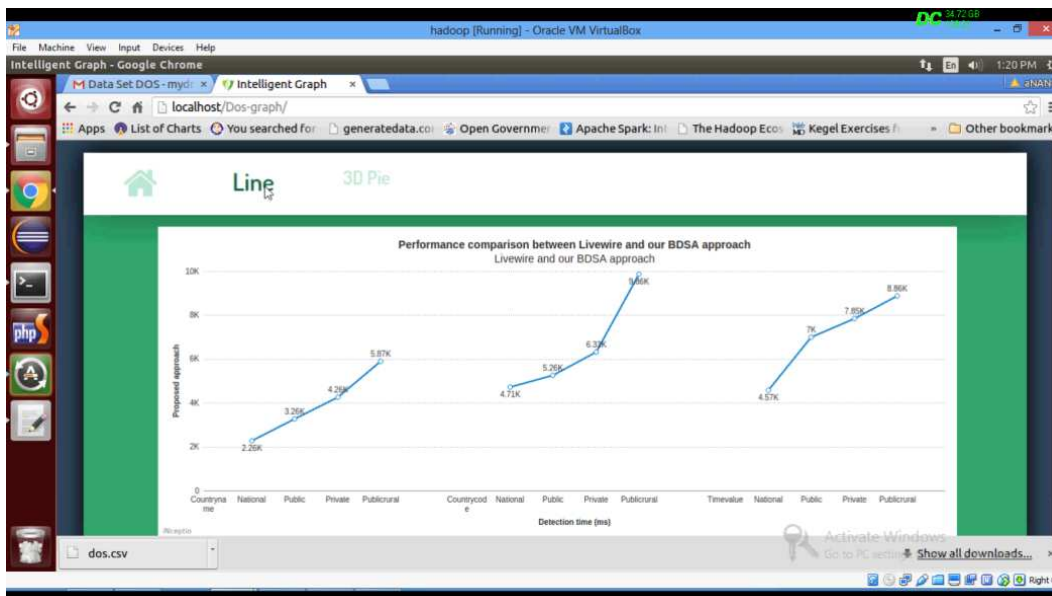


Fig.7: Performance comparison between Livewire and BDSA

V. CONCLUSION

The research paper presents a novel approach to protect virtualized infrastructures against advanced attacks. The BDSA approach first extracts attack features through graph-based event correlation, a MapReduce parser based identification of potential attack paths and then ascertains attack presence through two-step machine learning, namely logistic regression and belief propagation.

REFERENCES

- [1] B. Niloufer, S. Jessica Saritha, "Implementation of Big Data Protection Analytics in Virtualized Infrastructures", International Journal of Technical Innovation in Modern Engineering & Science (IJTIMES), Volume-4, Issue-9, 2018.
- [2] Dr. M. Bhanu Sridhar, A. Koushik, "A Study of Big Data Analytics in Clouds with a Security Perspective", International Journal of Engineering Research & Technology (IJERT), Volume-6 Issue-1, 2017.
- [3] Mr J L Aldo Stalin, M Badhri Narayanan, Mr. M Deepak, "Security Analytics for Protecting Virtualized Infrastructures", Information Systems & eBusiness Network. 2017.
- [4] Thu Yein Win, Huaglory Tianfield, Quentin Mair, "Big Data Based Security Analytics for Protecting Virtualized Infrastructures in Cloud Computing", IEEE, 2017.
- [5] Chaowei Yang, Qunying Huang, Zhenlong Li, Kai Liu, Fei Hu "Big Data and cloud computing: innovation opportunities and challenges", International Journal of Digital Earth, Volume-10, issue-1, 2017.
- [6] Motukuri Prashanthi, "Analysis of Security Issues in Virtualization Cloud Computing", International Journal of Computer Science and Mobile Computing IJCSMC, Volume-5, Issue-8, 2016.
- [7] Ali Gholami, Erwin Laure, "BIG DATA SECURITY AND PRIVACY ISSUES IN THE CLOUD", International Journal of Network Security & Its Applications (IJNSA) Volume-8, Issue-1, 2016.
- [8] Varsha, Amit Wadhwa, Swati Gupta, "Study of Security Issues in Cloud Computing", International Journal of Computer Science and Mobile Computing IJCSMC, Volume-4, Issue-6, 2015.
- [9] Lidong Wang, Cheryl Ann Alexander, "Big Data in Distributed Analytics, Cybersecurity, Cyber Warfare and Digital Forensics", Science and Education Publishing, Volume-1, Issue-1, 2015.
- [10] Monjur Ahmed, Mohammad Ashraf Hossain, "Cloud Computing and Security Issues In the Cloud", International Journal of Network Security & Its Applications (IJNSA), Volume-6, Issue-1, 2014.
- [11] Venkata Narasimha Inukollu, Sailaja Arsi, Srinivasa Rao Ravuri, "Security Issues Associated With Big Data in Cloud Computing", International Journal of Network Security & Its Applications (IJNSA), Volume-6, Issue-3, 2014.
- [12] T. Swathi, K.Srikanth, S.Raghunath Reddy, "Virtualization in Cloud Computing", International Journal of Computer Science and Mobile Computing IJCSMC, Volume-3, Issue-5, 2014.
- [13] Muhammad Kazim, Rahat Masood, Muhammad Awais Shibli, Abdul Ghafoor Abbasi, "Security Aspects of Virtualization in Cloud Computing", 2013.
- [14] Farzad Sabahi, "Secure Virtualization for Cloud Environment Using Hypervisor-based Technology", International Journal of Machine Learning and Computing, Volume-2, Issue-1, 2012.
- [15] Wentao Liu, "Research on Cloud Computing Security Problem and Strategy", IEEE, 2012.
- [16] Shubhashis Sengupta, Vikrant Kaulgud, Vibhu Saujanya Sharma, "Cloud Computing Security - Trends and Research Directions", IEEE, 2011.
- [17] Pedro Caldeira Neves, Bradley Schmerl, Jorge Bernardino, Javier Cámara, "Big Data in Cloud Computing: features and issues".