International Journal of Technology in Computer Science & Engineering
ISSN 2349 - 1582
Volume 4, No 2, June 2017

Development of Hybrid Intrusion Detection System on Big Data for Detecting Unknown attacks by using AHSVN

S. Ramesh[1]
Research Scholar,
Manonmaniam Sundaranar University,Tirunelveli, T.N, India
ramesh20062000@gmail.com

Dr. H. Abdul Rauf [2]
Dean & Professor, Department of CS & Technology
Sri Sastha Institute of Engineering and Technology, Chennai, T.N, India
harauf@yahoo.com

Dr. S.P. Victor [3]
Dean of Science & Associate Professor of Computer Science
St. Xavier's College, Palayamkottai, T.N, India
drspvictor@gmail.com

Abstract: Big data is a growing term that describes any voluminous amount of structured, semi structured and unstructured data that has the potential to be quarried for information. Big data is a word for data sets that are so large or complex that traditional data processing applications are not enough to deal with them. Challenges in big data include analysis, capture, data curation, search, sharing, storage, transfer, visualization, querying, updating and information privacy. As for as the security of Big data is concerned the volume of data along with the bandwidth speed it generates makes it difficult for handling big data with currently available security tools. The existing research concentrates on malicious attacks only. Network hacking crime is going on increasing since the available security technologies are not performed well in big data environment. Due to the increasing cyber attacks, constructing Hybrid intrusion detection system is necessary for preventing from attacks. But the multiple IDS approach to co-exist in a single system is a major challenging task. To overcome this issue, we propose an Adaptive Hybrid Support Vector Network Algorithm (AHSVN). This mechanism used in the Intrusion Detection scheme helps to identify the attack at the early stage and give alarms and also the node will generate alerts to other nodes if there is a possibility of attack. We have tested the viability of our AHSVN Algorithm by conducting experiments. The experimental results show that our proposed Algorithm can give efficient Intrusion detection system for unknown attacks in Big data environment.
Keywords: AHSVN

1. Introduction

Telecommunication networks are getting more important in our social lives because many people want to share their information and ideas. Thanks to the rapid development of the Internet and ubiquitous technologies including mobile devices such as smart phones, mobile phones and tablet PCs, the quality of our lives has been greatly influenced and rapidly changed in recent years. Internet users have exponentially increased as well. Meanwhile, the explosive growth of teletraffic called big data for user services threatens the current networks, and we face menaces from various kinds of intrusive incidents through the Internet. A variety of network attacks on network resources have continuously caused serious damage. [1]

Organizations are paying huge amount only for securing their confidential data from attackers or intruders. But the hackers are Big Bosses and are very sharp enough to crack the security of the organization. Therefore before they made security breach, let us hunt down them and make the alert for organization, so that they can save their

confidential data. For the above mentioned purpose, Intrusion detection system came into existence. But the current systems are not capable enough to detect all the attacks coming towards them.[2]

Recently, threats of previously unknown cyber-attacks are increasing because existing security systems are not able to detect them. Past cyber-attacks had simple purposes of leaking personal information by attacking the PC or destroying the system. However, the goal of recent hacking attacks has changed from leaking information and destruction of services to attacking large-scale systems such as critical infrastructures and state agencies. In the other words, existing defence technologies to counter these attacks are based on pattern matching methods which are very limited. Because of this fact, in the event of new and previously unknown attacks, detection rate becomes very low and false negative increases.[3]

A hybrid approach of intrusion detection through knowledge discovery from big data using Latent Dirichlet Allocation (LDA). To identify the "hidden" patterns of operations conducted by both normal users and malicious users from a large volume of network/systems logs, by mapping this problem to the topic modelling problem and leveraging the well established LDA models and learning algorithms.[4]

With the rapid development of Internet and Information technology, detecting network intrusion behaviours have been attracted more and more attentions. The novel intrusion detection algorithm using a hybrid ant colony and support vector machine model. The framework of the detecting network intrusion system is given which is designed to promote the accuracy of detecting network intrusion by optimizing parameters of support vector machine with ant colony algorithm.[5]

The rate of data generation is enormously growing due to the number of internet users and its speed. This increases the possibility of intrusions causing serious financial damage. Detecting the intruders in such high-speed data networks is a challenging task. So high-speed Intrusion Detection System (IDS), capable of working in Big Data environment. The system design contains four layers, consisting of capturing layer, filtration and load balancing layer, processing layer, and the decision-making layer..[6]

2. Literature Survey

In 2012, **Hae-Duck J.Jeong, WooSeok Hyun,** et al. proposed an anomaly tele traffic intrusion detection system based on the open source software platform Hadoop, and some problems and solutions for this system have been also investigated. The proposed framework will be developed and experimented with on Hadoop in the future.

In 2013, **Alvaro A.Cardenas,Pratyusa K.Manadhata,** et al. concluded that Big data is changing the landscape of security technologies for network monitoring and forensics. However, in the eternal arms race of attack and defence, big data is not a solution, and security researchers must keep exploring novel ways to contain sophisticated attackers. Big data can also create a world where maintaining control over the revelation of our personal information is constantly challenged. This paper give importance to increase the efforts to educate a new generation of computer scientists and engineers on the value of privacy and work with them to develop the tools for designing big data systems..

In 2014, **Bilal Maqbool Beigh** et al. proposed a new hybrid model which combines both the modules i.e anomaly and signature based also it added few more modules which will check for false alarm and will add new rules to the signature database automatically.

In 2014, **Sung-Hwan Ahn, Nam-Uk Kim** et al. proposed big data system model for reacting to previously unknown cyber threats and worked on the deduction of practical technologies. The proposed technique will extract information from a variety of sources to detect future attacks..

In 2014, **Vandana P.Janeja, Ali Azari** et al. concluded that they have proposed a Big-distributed Intrusion Detection System. In this architecture, using HAMR, a big data processing engine and proposed a novel ensemble method to identify multi-pronged attacks

In 2015, **Hu Jianhong** proposed a new hybrid model which combines both the modules i.e anomaly and signature based also added few more modules which will check for false alarm and will add new rules to the signature database automatically. This model will detects novel intrusion made on the system and is reducing the number of false alarm.

In 2016, **M. Mazhar Rathore** et al. proposed a novel real-time intrusion detection system that can work in a high-speed network environment. The system includes the proposed architecture with four-layers, the parameters selection mechanism and the proposed intrusion detection technique. The processing layer, which is the main component of the system, is composed of various Hadoop master and data nodes, which is responsible for handling high speed real-time traffic more efficiently in order to identify any type of intrusions in the network

3. Objective

The Network security plays a key role for many organizations and corporate. Generally host based and network based Intrusion Detection Systems are available in the market depending upon the detection technology used by them. The main objective of this research paper is maintaining security across the mixed data from uniform sources and co-relating the mixed data from different sources using hybrid strategy. A real time IP base master to client communication Intrusion Detection Systems (IDS) prevents security intrusions by gathering and composing with technologies. Due to the increasing cyber attacks, constructing Hybrid intrusion detection system is essential for preventing from attacks. Multiple IDS approach to co-exist in a single system is a major challenging task. To eliminate these issues we propose an Adaptive Hybrid Support Vector Network Algorithm (AHSVN).

4. Proposed System

The cyber attacks go on increasing due to the existing IDS technologies which are not capable of detecting the intrusions. Our proposed work will enrich the efficiency of the IDS as compare to the past IDS systems. The main work of this research papers is maintaining security across the external fraud detection data from homogeneous sources and co-relating the heterogeneous data from different sources using hybrid strategy over Big Data. The proposed system specifies a set of IP base rules AHSVN in Hadoop concept one can attempt to get good results by improving the efficiency and reducing the complexity present in the model of external unknown attacks.

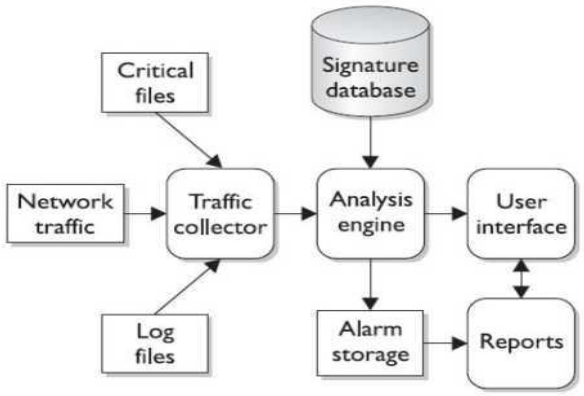

Fig:1 Intrusion Detection process

The foundation for a layered network security approach usually starts with a well secured system. A well-secured system uses up-to-date application and operating system patches, well-chosen passwords, the minimum number of services running, and restricted access to available services. On top of that foundation, could add layers of protective measures such as antivirus products, firewalls, sniffers, and IDSs. Some of the more complicated and interesting types of network/data security devices are IDSs, which are to the network world what burglar alarms are to the physical world. The main purpose of an IDS is to identify suspicious or malicious activity, note activity that deviates from normal behaviour, catalogue and classify the activity, and if possible, respond to the activity.

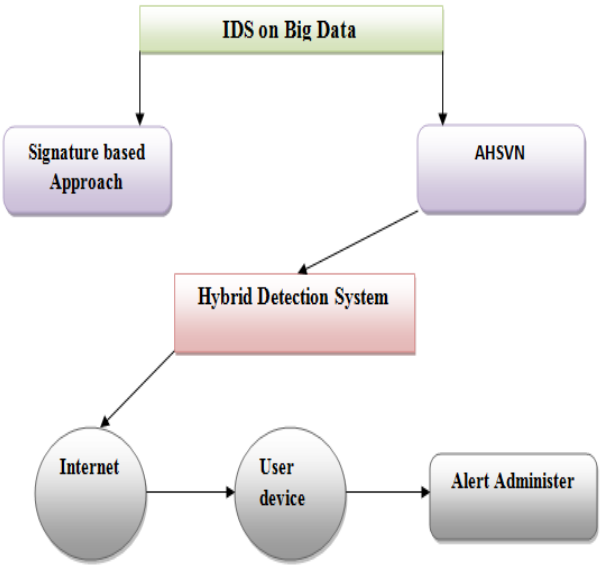

Fig: 2 Proposed Block IDS diagram

The Block diagram of proposed system is shown in Fig: 2. Signature based approach block finds out the intrusion activity in existing packet with the help of attack rules and if found then applies appropriate rules otherwise it drops the packet. It takes different time to respond different packets that depends upon the power of machine and number of rules defining the system. AHSVN block has detection engine, that might generate an alert or used to do log activities. Depending upon the nature of data all log files are kept by default in a folder and by using command line options the location can be changed.

The output modules saves output generated by the logging and alerting system of AHSVN. The administrator will get the alert from the IDS by database searching and query building by AHSVN. Searches can be performed between network specific parameters such as the attackers internet protocol address, and other parameters such as time or date of an event, by AHSVN triggered rules.

5. Proposed Algorithm AHSVN

Advance Pattern analysis includes categorization which enclose support vector network algorithm. Support Vector Network is clustering algorithm which is used for discovery of network attacks. Support Vector Network is second-hand for intrusion detection system to notice which IP has viruses and then clean that IP packet.

The IP based Hybrid support vector network is a new learning machine algorithm. In Hybrid Support Vector Network algorithm, data set is divided into preparation part and testing part. The existing methodology has high error rate. So the AHSVM classify the attacks using support vector network learning algorithm. The major reasons for using AHSVM in IDS are speed. AHSVN can learn a larger set of patterns and can scale better than any existing methods. Feature selection or attribute reduction can help to reduce the SVM classification time and saving memory space effectively and efficiently.

1. Capture network data.
2. D ← Stored data from database.
3. N ← all feature set.
4. th← threshold value.
5. **for** $i = 1 \rightarrow n$ **do**
6. F = F - Fi
7. ac = calculate Accuracy(F)
8. **if** ac ≤ th **then**
9. break
10. **end if**
11. **end for**
12. Apply learning algorithms.
13. Classified Data.
14. End

6. Design and Implementation

AHSVN algorithm for categorization: "Support Vector Network" (SVN) is a supervise machine learning algorithm which can be used for detecting intrusions.

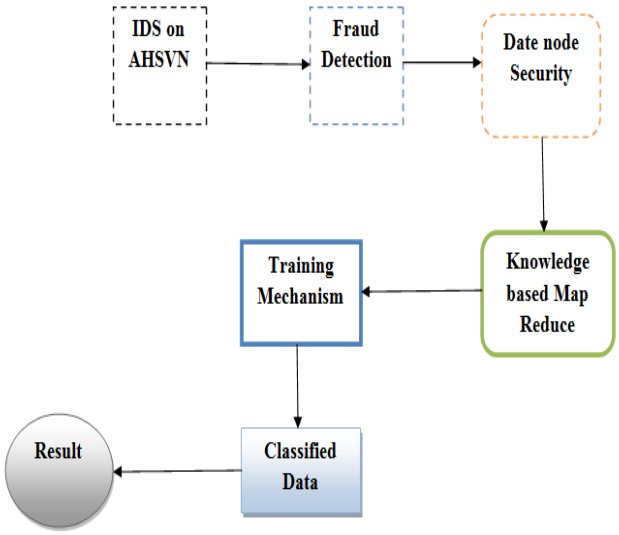

Fig: 3 Shows the Block diagram of Proposed AHSVN System

The AHSVN operations stack which includes application, systems software, infrastructure, network etc. Fraud detection is getting easy due to the event collection and logging system. The Real-time data IP analysis works well when it is based on predefined algorithms and queries in hybrid training mechanism. Current tools are able to synthesize multiple streams of rapidly flowing data and perform complex operations on them. At every layer of the system's hardware and application stack, real-time analytics enable centralized log collection and monitoring system. These processes are used to find normal and abnormal IP flow.

7. Experimental Results

The results obtained in the proposed system are given in the following table and graph.

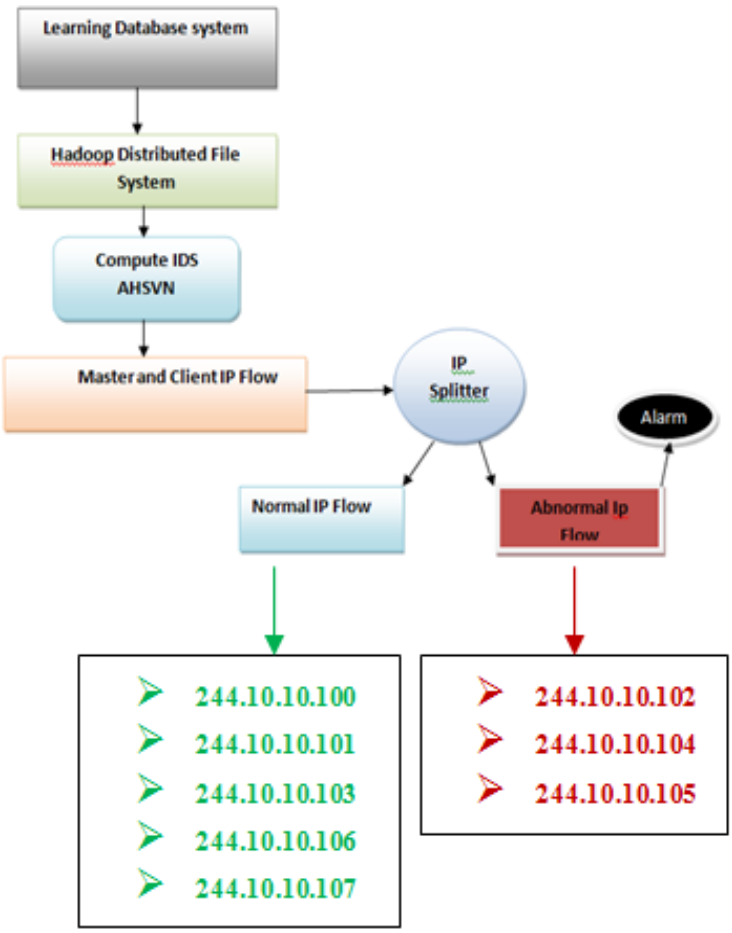

Fig: 4 IDS in Big Data system using IP's

Time(ms)	Bytes	Normal IP IDS	Abnormal IP IDS
12:33:110	76	244.10.10.100	
12:34:112	76	244.10.10.101	
12:36:114	76		244.10.10.102
12:38:115	76	244.10.10.103	
12:39:116	76		244.10.10.104
12:40:117	76		244.10.10.105
12:41:118	76	244.10.10.106	
12:42:119	76	244.10.10.107	

Table: 1 Intrusion identification based on IP's

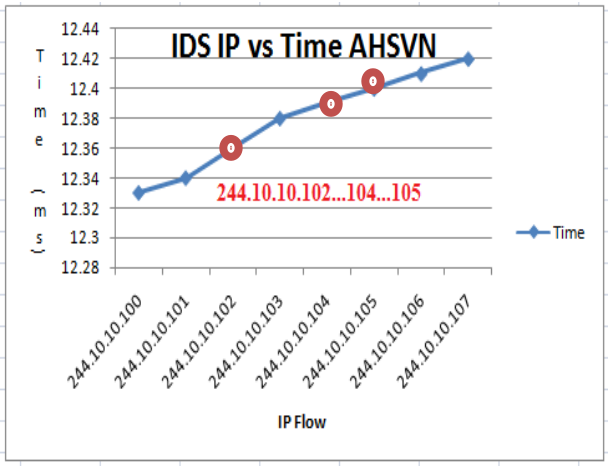

Fig 5. Graph Analysis Flow over IDS

8. Conclusion

To protect from various threats in Big Data analysis system, data sets have been analysed with our new technique, IP based Adaptive Hybrid Support Vector Network algorithm to prevent external unknown attacks.

In this paper, we have obtained the information of external fraud detection attack a new move towards external fraud Detection approach to notice the intrusion in the big data analysis network. Our training model consists of big datasets with dispersed environment that improve the performance of Intrusion detection system. The proposed approach mainly focuses on unknown external fraud detection in big data, using Adapive Hybrid Support Vector Network algorithm(AHSVN). The experiment showed that using the proposed method better results were obtained. This shows the efficiency of the proposed method. When Abnormal IP enters in to the IDS the IDS will filter the abnormal packets and give alerts.

9. Future work

In future a real time Intrusion Detection system for Big Data system using AHSVN algorithm may be developed in such a way to give improved results.

10. Reference

[1] Hae-Duck J. Jeong, WooSeok Hyun, Jiyoung Lim, and Ilsun You “Anomaly Teletraffic Intrusion Detection Systems on Hadoop-Based Platforms: A Survey of Some Problems and Solutions” 2012 15th International Conference on Network-Based Information Systems Pages: 766 - 770, DOI: 10.1109/NBiS.2012.139

[2]. Bilal Maqbool Beigh “One-stop: A novel hybrid model for intrusion detection system” 2014 International Conference on Computing for Sustainable Global Development (INDIACom) Pages: 798 - 805, DOI: 10.1109/IndiaCom.2014.6828072

[3] Sung-Hwan Ahn, Nam-Uk Kim, and Tai-Myoung Chung “Big data analysis system concept for detecting unknown attacks” 16th International Conference on Advanced Communication Technology Pages: 269 - 272, DOI: 10.1109/ICACT.2014.6778962

[4]. Jingwei Huang, Zbigniew Kalbarczyk, and David M. Nicol “Knowledge Discovery from Big Data for Intrusion Detection Using LDA” 2014 IEEE International Congress on Big Data Pages: 760 - 761, DOI: 10.1109/BigData.Congress.2014.111

[5]. Hu Jianhong “Network Intrusion Detection Algorithm Based on Improved Support Vector Machine” 2015 International Conference on Intelligent Transportation, Big Data and Smart City Pages: 523 - 526, DOI: 10.1109/ICITBS.2015.135

[6]. M. Mazhar Rathore, Anand Paul, Awais Ahmad, Seungmin Rho, Muhammad Imran, and Mohsen Guizani “Hadoop Based Real-Time Intrusion Detection for High-Speed Networks” 2016 IEEE Global Communications Conference (GLOBECOM) Pages: 1 - 6, DOI: 10.1109/GLOCOM.2016.7841864

[7]. Mostafa Doroudian, Narges Arastouie, Mohammad Talebi, and Ali Reza Ghanbarian “Multilayered database intrusion detection system for detecting malicious behaviors in big data transaction” 2015 Second International Conference on Information Security and Cyber Forensics (InfoSec) Pages: 105 - 110, DOI: 10.1109/InfoSec.2015.7435514

[8]. Chaitali Gupta, Ranjan Sinha, and Yong Zhang “Eagle: User profile-based anomaly detection for securing Hadoop clusters” 2015 IEEE International Conference on Big Data (Big Data) Pages: 1336 - 1343, DOI: 10.1109/BigData.2015.7363892

[9]. Alvaro A. Cárdenas, Pratyusa K. Manadhata, and Sreeranga P. Rajan “Big Data Analytics for Security” IEEE Security & Privacy. Volume: 11, Issue: 6 Pages: 74 - 76, DOI: 10.1109/MSP.2013.138

[10]. Vandana P. Janeja, Ali Azari, Josephine M. Namayanja, and Brian Heilig “B-dids: Mining anomalies in a Big-distributed Intrusion Detection System” 2014 IEEE International Conference on Big Data (Big Data) Pages: 32 - 34, DOI: 10.1109/BigData.2014.7004484

[11]. Manoj Kumar, Robin Mathur “Unsupervised outlier detection technique for intrusion detection in cloud computing” International Conference for Convergence for Technology-2014 Pages: 1 - 4, DOI: 10.1109/I2CT.2014.7092027

[12]. Zhiyuan Tan, Upasana T. Nagar, Xiangjian He, Priyadarsi Nanda, Ren Ping Liu, Song Wang, and Jiankun Hu “Enhancing Big Data Security with Collaborative Intrusion Detection” IEEE Cloud Computing Volume: 1, Issue: 3 Pages: 27 - 33, DOI: 10.1109/MCC.2014.53

[13]. Samuel Marchal, Xiuyan Jiang, Radu State, Thomas Engel “A Big Data Architecture for Large Scale Security Monitoring” 2014 IEEE International Congress on Big Data Pages: 56 - 63, DOI: 10.1109/BigData.Congress.2014.18

[14]. Mostafa Doroudian, Narges Arastouie, Mohammad Talebi, Ali Reza Ghanbarian “Multilayered database intrusion detection system for detecting malicious behaviors in big data transaction” 2015 Second International Conference on Information Security and Cyber Forensics (InfoSec) Pages: 105 - 110, DOI: 10.1109/InfoSec.2015.7435514

[15]. Kalpana Jaswal, Praveen Kumar, Seema Rawat “Design and development of a prototype application for intrusion detection using data mining” 2015 4th International Conference on Reliability, Infocom Technologies and Optimization (ICRITO) (Trends and Future Directions) Pages: 1 - 6, DOI: 10.1109/ICRITO.2015.7359266

[16]. Goutam Mylavarapu, Johnson Thomas, and Ashwin Kumar TK “Real-Time Hybrid Intrusion Detection System Using Apache Storm” 2015 IEEE 17th International Conference on High Performance Computing and Communications, 2015 IEEE 7th International Symposium on Cyberspace Safety and Security, and 2015 IEEE 12th International Conference on Embedded Software and Systems Pages: 1436 - 1441, DOI: 10.1109/HPCC-CSS-ICESS.2015.241

[17]. Xiaoming Zhang, Guang Wang “Hadoop-Based System Design for Website Intrusion Detection and Analysis” 2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity) Pages: 1171 - 1174, DOI: 10.1109/SmartCity.2015.231

[18]. Zhiyuan Tan, Upasana T. Nagar, Xiangjian He, Priyadarsi Nanda, Ren Ping Liu, Song Wang, and Jiankun Hu “Enhancing Big Data Security with Collaborative Intrusion Detection” IEEE Cloud Computing 2014, Volume: 1, Issue: 3 Pages: 27 - 33, DOI: 10.1109/MCC.2014.53

[19]. Sanchita Arora, Mithun Kumar, Prashant Johri, Sanjoy Das "Big heterogeneous data and its security: A survey” 2016 International Conference on Computing, Communication and Automation (ICCCA) Pages: 37 - 40, DOI: 10.1109/CCAA.2016.7813727

[20]. Viktoriya Degeler, Richard French, Kevin Jones “Self-Healing Intrusion Detection System Concept” 2016 IEEE 2nd International Conference on Big Data Security on Cloud (BigDataSecurity), IEEE International Conference on High Performance and Smart Computing (HPSC), and IEEE International Conference on Intelligent Data and Security (IDS) Pages: 351 - 356, DOI: 10.1109/BigDataSecurity-HPSC-IDS.2016.27

[21]. Cuong Tuan Nguyen; Masaki Nakagawa Finite State Machine Based Decoding of Handwritten Text Using Recurrent Neural Networks 2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR) Pages: 246 - 251, DOI:10.1109/ICFHR.2016. 0055

[22]. Shengyi Pan; Thomas Morris; Uttam Adhikari Developing a Hybrid Intrusion Detection System Using Data Mining for Power Systems IEEE Transactions on Smart Grid Volume: 6, Issue: 6 Pages: 3104 - 3113, DOI: 10.1109/TSG.2015.2409775

[23]. Santosh Aditham; Nagarajan Ranganathan; Srinivas Katkoori Memory access pattern based insider threat detection in big data systems 2016 IEEE International Conference on Big Data (Big Data) Pages: 3625 - 3628, DOI: 10.1109/BigData. 2016.7841027

[24]. Rajesh Sankaran; Ricado A. Calix On the feasibility of an embedded machine learning processor for intrusion detection 2016 IEEE International Conference on Big Data (Big Data) Pages: 1082 - 1089, DOI: 10.1109/BigData.2016.7840711

[25]. Saad Mohamed Ali Mohamed Gadal; Rania A. Mokhtar Anomaly detection approach using hybrid algorithm of data mining technique 2017 International Conference on Communication, Control, Computing and Electronics Engineering (ICCCCEE) Pages: 1 - 6, DOI: 10.1109/ICCCCEE.2017.7867661

[26]. Jingwei Huang; Zbigniew Kalbarczyk; David M. Nicol Knowledge Discovery from Big Data for Intrusion Detection Using LDA 2014 IEEE International Congress on Big Data Pages: 760 - 761, DOI: 10.1109/BigData.Congress.2014.111

[27]. Shin-Ying Huang; Yennun Huang; Neeraj Suri Event Pattern Discovery on IDS Traces of Cloud Services 2014 IEEE Fourth International Conference on Big Data and Cloud Computing Pages: 25 - 32, DOI: 10.1109/BDCloud.2014.92

[28]. Florian Gottwalt; Achim P. Karduck SIM in light of big data 2015 11th International Conference on Innovations in Information Technology (IIT) Pages: 326 - 331, DOI: 10.1109/INNOVATIONS.2015.7381562

[29]. Miss Gurpreet Kaur Jangla1, MrsDeepa. A.Amne Development of an Intrusion Detection System based on Big Data for Detecting Unknown Attacks 2015 International Journal of Advanced Research in Computer and Communication Engineering Vol. 4, Issue 12

[30]. Rupali V.Molawade Vijaya S.Waghmare Big Heterogeneous Data for Intrusion Detection 2015 International Journal of Computer Applications(0975–8887)International onference on Advances in Science and Technology 2015