

Analysis of How Scalable Features in Hadoop/MapReduce by Internet Traffic Management

Deden Witarsyah Jacob^a, Muhammed.E Abd Alkhalec Tharwat^{b,*}, Mohd Farhan Md Fudzee^b,
Azizul Azhar Ramli^b, Shahreen Kasim^b, Muharman Lubis^a

^a Department Information System, Telkom University, Jl. Telekomunikasi, Bandung, 40257, Indonesia

^b Faculty of Computer Science and Information Technology, Universiti Tun Hussein Onn Malaysia, 86400, Malaysia

Corresponding author: *muhammead@gmail.com; gi170023@siswa.uthm.edu.my

Abstract— Internet traffic monitoring is to measure and analyze the network bottlenecks to manage the online data are transferring processes efficiently. Various tools have been developed by using internet traffic measurement and internet traffic analysis tools, such as Hadoop. Activity measurement and adaptive examination represent the dynamics of information exchange. On the other hand, information exchange and dynamics measure movement in light of the system assets that can be accessed depending on the characteristics of the exchanged information. The main aim of this work is to apply scalable features of internet traffic measurement and analysis using Hadoop to understand the effects of these features on the speed of transferring data. This gives a new vision or opportunity to dynamically adapting the most suitable traffic measurement and analysis feature according to network capabilities and environment. This research employs Hadoop/Map Reduce as scalable internet traffic measurement and analysis tools. The simulation was conducted by using five personal computers; one as a server and four virtual computers as network nodes. Each computer has 2GB memory and 100GB storage. Five types of data segmentation are utilized 10 MB, 40MB, 64MB, 200MB, and 500MB. The speed of the network is calculating in a megabit per second (Mbs) based upon the network speed on the number of allocated PCs (100 Mbs/4). The simulation is conducted to test the data transfer time based on various selections of network capabilities such as transferring extensive data through a network of medium and heavy usage.

Keywords— Internet traffic measurement; traffic network; data transfer and sharing; Hadoop; Map Reduce.

Manuscript received 21 Jan. 2020; revised 17 May 2020; accepted 16 Oct. 2020. Date of publication 28 Feb. 2023.
IJASEIT is licensed under a Creative Commons Attribution-Share Alike 4.0 International License.



I. INTRODUCTION

The scalability of internet traffic measurement and analysis, which are affected by user behaviors and network users, need to be checked and reviewed necessarily due to the significant increase in recent data. Internet traffic measurement and analysis become a difficult task due to the need for addressing the large scales data set and matching computing and storage resources [1]. Various tools have been developed based on internet traffic measurements and internet traffic analysis tools like Tcpdump, Coral Reef, Snort, Tstat, 7 Filter, and NetFlow. For easier data compile and analysis, we need a tool to map the Distributed File System, which is helping to manage the regulating traffic. The Hadoop tool is introduced as an alternative for that purpose. It is distributed file systems that consist of an open-source computing program (MapReduce), which is the proper infrastructure for analyzing big data. Additionally,

easy access to storage services is achieved with the help of a distributed computing system while accommodates the expansion of data that are processed [2].

Internet traffic measurement and analysis must detect the problem of sending data by active networks like data losses and slow the information transfer. In Hadoop/Map Reduce, there are two main process steps. First, start with measuring network traffic to hold and segment the data into a packet, and second analyze network traffic packets that pass through available paths based on the User Datagram Protocol (UDP) and Transmission Control Protocol (TCP), which are the main protocols of the network layer.

However, the processes of traffic measurement and analysis in applications such as Hadoop are accomplished based on fixed-configuration, whatever the changes of the network environment. In this study, the simulation was applied to two main transferring protocols approaches that are TCP and UDP. The significant results of the simulation test show that the TCP is more effective for transferring

protocol as compared to the UDP approach. The network measurement and analysis using Hadoop/Map Reduce based on fixed transferring options are not recommended because the speed of data transfer is effected by transferring option such as the level of network traffic, data segmentation, and the type of transfer protocol. The Hadoop/Map Reduce should be deployed in a dynamic environment of network traffic. The best features of measurement and analysis processes should be selected automatically based on the network traffic capabilities [3].

II. MATERIALS AND METHOD

A. Hadoop Overview

Hadoop is one of the open-sourced platforms utilized to handle and process the immense size of data. Classical companies widely build based on costly hardware devices to store and to process their data. However, Hadoop comes to the picture to fulfill the company requirement to store and processing vast amounts of data across existing servers. Hadoop has many features that make it applicable to use, such as, firstly, it can be built and developed using any programming language such as (C++, JAVA). Secondly, it is compatible with any operating system like Windows and Linux. Third, it can handle traffic measurement and analysis as a single packet. And finally, it can update traffic measurement functions such as (scaling network usage levels and providing different data segments) [4].

With Hadoop implementation, data processing will be faster and efficient with the help of adding new hardware devices without limits. This is a benefit because companies need to collect data from new resources daily. The ability allows organizations to preserve the value of data that was already considered as not useful [5]. There is various Internet-related work, but it is restricted to some Hadoop data without full coverage and gives you the ability to track and measure this significant data with MapReduce [6]. A previous study recommends some of the tips of Hadoop cluster optimization after the analyses of the existing problems in the Hadoop data processing platform [7]. This work also describes the architecture of the Hadoop cloud computing model and its application status. In [8] the authors designed a network traffic prediction system based on Hadoop to process real mobile network traffic data for a major network operator in China. Through Echo State Network (ESN), the system provides an efficient way to handle large-scale network records and build predictive models at the same time. It has been analyzed by the previous study the best Hadoop system by improving the MapReduce parallelization framework and improving the performance of Hadoop Distributed File System (HDFS) [9].

Hadoop Features consist of five aspects, as follows:

- It is open-source.
- It is a distributed storage
- Using the MapReduce can process the big data
- Reliable and easy to use.
- References locality

The components of the Hadoop are listed below. Figure 1. represents the Hadoop components framework.

Hadoop Distributed File System (HDFS). It is the backbone of the Hadoop, which provide reliable and

efficient cost storage for big data

Distributed Processing (MapReduce): It represents the core of Hadoop, which is responsible for processing the data in parallel and helps to improve the speed and reliability of the cluster. It is breaking the work into 2 phases (Map and Reduce).

Operating System (YARN): It's one of the essential components in Hadoop, which is known as the operating system, which helps to manage and monitor the workloads. YARN is flexible, efficient, and share.

Warehouse (Hive): It's the data warehousing system utilized for analyzing the large datasets, queue it in the Hadoop file, and finally provide data summarization. Hive has four parts (Metastore, Driver, Query compiler, and Hive server).

Data Science Tools (Mahout): It's a machine learning and data mining framework. It has many algorithms like clustering, classification, and collaborative filtering. It provides meaning to the data after stored in HDFS.

Analyzing and Querying (Pig): It's another component of Hadoop which utilized to analyze extensive scale data that are stored in the Hadoop distributed file system (HDFS). A pig can handle any type of data using high-level language PigLatin.

Management Platform (Ambari Component): It is a consistent management platform utilized to manage and monitor the Hadoop cluster. Ambari is secure, easy to modify and customize it, and simple to install and configure.

Store Structured Data(HBase): It is a distributed database that stores the data in a table with a huge number of rows and columns in real-time access that can easily write/read in HDFS.

Imports and Exports Data (Sqoop): It is an import and export data from the Hadoop to outside sources. It is can work in the parallel transfer, which efficiently led to fast copy.

Workflow and Scheduling (Oozie): It is a Hadoop framework component that is responsible for merging many jobs in one logic unit I sequence. There are two types of Oozie workflow and coordinator. The first one is stored and runs the workflows of jobs (tasks). The second one, it's run the workflow based on the predefined schedules.

Maintaining Configuration (Zookeeper): One of the fast component in Hadoop which represent the center service to retain and provide group services.

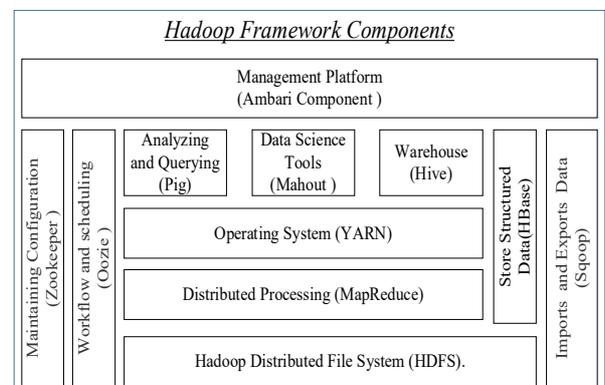


Fig. 1 Hadoop Components

B. The Significance of Network Observation Alarms

Through the last decennium or is such a thing, Investment, the Internet among individuals, and quickly expanded from going about as a means vital that composite correspondence. The internet information exchange and participation likewise developed quickly through numerous Internet applications, for example, messages, interpersonal organizations; interactive media or video administrations, and talks. Therefore, Internet use additionally has experienced substantial changes. Thus, the multiple benefits of the Internet prompt increment in the Internet system movement. According to the Internet, traffic brings about an array of problems [10].

- Delay of data transfer and sharing.
- Loss of transferred are data due to network routing.
- The increased cost of the network requirements is due to congestion in the traffic.
- Increased opportunities for an attacker to steal or manipulate online data.

To avoid Internet traffic, new protocols, routing algorithms, and tools have been developed. Mainly that approaching and protocols target in measuring and it analyses the network obstacles in management the internet data conversion process efficiently. However, notwithstanding those protocols, Internet traffic measurement and analysis still confront several challenges. A large data set requires a transfer at real-time and accurately, i.e., affective computing based on network fixed assists such as storage resources.

C. Internet Traffic

Web correspondence associates a large number of Internet clients over the world. Moore than four billions of users are connecting across the globe using the Internet. The internet system has a considerable infrastructure based on optical fiber, a chain of copper wires, and wireless connections. With the help of devices and applications such as a personal computer, mobile cell, routers, switches, and some small devices that are typically used in homes, in cars, and daily schedule. On the other hand, there is much application utilized to support the Internet, such as the web browser, e-mail system, sharing and transfer data system, video conference, chatting applications, games, and many more. As a considerable base framework, the Internet additionally bolsters a variety of utilization.

The applications incorporate global Internet, messages, registry sharing, telephony, radio, computer games," and other business services." And from an essential inquiry to deliver according to how about active or portray the online movement effectively. Web activity portrays concerning when and how the movement might be examined. When contrasted with a bigger organization, the practices of the movement system are very impressive by the activity practices in a little organization. It might likewise be noticed which, with the new orders, the qualities with conduct from the activity and can be changed, for example, a regulation which is about new species. An overview led measured that around 70 to 75% of the movement was contained web activity. From there forward, we have seen a huge expansion in the aggregate activity volumes; web activity still involves a bigger offer of numerous systems.

Recently, the sharing of the document regularly rules the

movement application. Besides, TV and video conveyance over IP has turned out to be generally spread, and create expanding activity volumes. In this manner, online movement depends on system assets, and information exchanged through these roles. Exchange rapid at the same time. For example, exchanged to store the information prompt Internet activity under-resourced systems when the vast system asset, the crowded movement it may not happen [11]. The traffic of the Internet can lead to delay the data sharing and to transfer. Monitoring the network has many benefits, some of it is listed below:

- Gives a clear picture of the current situation of network performance and traffics.
- Provide a secure network based on low cost.
- Support to achieve the Service Level Agreement (SLA) of the organization [12].
- With network monitoring debugging and solving the errors become a faster and easy task.
- Help to identify the failure that could lead to deadlock the system.

D. Internet Traffic Measurement and Analysis

These days, the online developed according to the main applications used broadly for the benefit of individual, authority, and business correspondence. One of the main variables which contribute to the growth of sensational development and expansion of the Internet is its perfect qualities, for example, adaptability. (The adaptability and flexibility of the online emerged by connecting any electronic or advanced gadget to the Internet, which may extend from routine desktop/PCs to supercomputers or bigger servers covering numerous sorts of wires gadgets), for example, hand telephones, sensors. Also, we can witness unprecedented changes in the use of the Internet differs from that of the prior use in 1969 [13].

At this time, the Internet is the essential key for transmitting data across the world, and data can be presented in text, video, or images. It helps connect users from all over the world with the help of various methods. It is possible that notes that one of the obstacles is rapid development, which does not leave adequate time and assets to absorb the estimation outcomes within internet framework, applications, and conventions. We can understand the system foundation and individual patterns after testing them in isolated lab situations and system reproductions, even though a worldwide scale hostile internet environment is not clear [14].

The difficulties in growing full comprehension increased for the reason that the shape of the Internet was not arranged ahead of time, where various systems of independent associations have been associated with the entire Internet. Therefore, we observe that the convention and utilization of internet change and develop with time c. We may describe the Internet as the essential correspondence mean. However, there are still critical inquiries about their capacities and functions. Besides, we can find a lot of network users unable to present acceptable answers regarding their network traffic due to the peer-to-peer (p2p) instructions of file sharing and shown in figure 2 application proposed method in the flowchart [15], [16]. Measurement of Internet traffic is highly demanded to fully understanding the internet scenario on highly aggregated connections. It is a complicated task to

measure internet traffic due to the challenges of it. Many tools have been proposed to measure and monitor the traffic of the Internet from The previous two decades, such as Tcpdump, CoralReef, Snort, Tstat, L7Filter, NetFlow, and Hadoop/MapReduce. In sub-section F has more details of each tool individually.

E. Flowchart of Proposed Method

In figure 2 the proposed method starts with the loading data and using the measurement and analysis tools. Based on Hadoop/Map Reduce (TCP, UDP) can measure the transfer data and check whether the transfer data are larger or less than 11.1. If the transferred data ≤ 11.1 than save to scenarios transfer data on TCP protocol and finally end. If the transferred data > 11.1 than, repeat to read new large data to measure it again. The main target of this study is to design a dynamic framework of Internet traffic measurement and analysis based on scalable transfer options.

F. Items of Internet Traffic Measured and Analyzed

The section lists the diagram from seven-famed Internet movement estimation with examination devices in interpreting focuses and develop the plan of these tools used. As delineated in figure 3, the main traffic measurement and analysis tools are the following: Hadoop /Map Reduce, Tcpdump, Coral Reef, Snort, Tstat, Net Flow, and L7 Filter [4], [15].

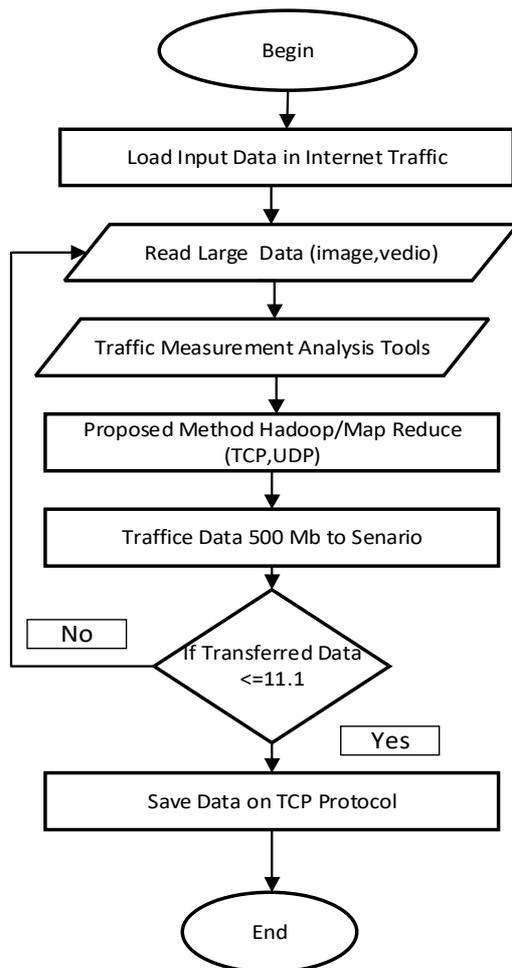


Fig. 2 Flowchart of The Proposed Method

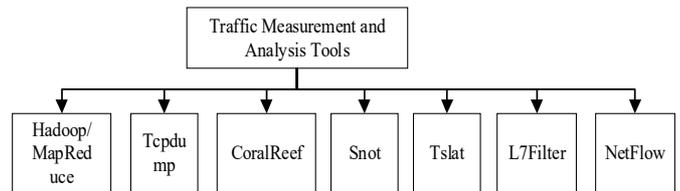


Fig. 3 Traffic Measurement and Analysis Tools

1) *Hadoop/MapReduce*: It is a software framework that is commonly utilized for a simple and easy application like writing. Hadoop/MapReduce is an authoritative tool that can address a large cluster of hardware as well as process a massive amount of data simultaneously [16]. Generally, MapReduce responsible for partitioning the input into free sets and process several tasks in a parallel performance manner. Usually, the filing system utilizes to store the input and output data. The Hadoop/MapReduce has various functions like monitoring, scheduling, and performing the failure of tasks. Commonly, the distributed file system of Hadoop (HDFS) and the MapReduce framework is run on similar sets of nodes due to the similarity between the storage nodes and computers [17]. This helps the framework to work effectively on the node when data already exists. Additionally, to implement the MapReduce applications, two utilities can be used: Hadoop Streaming and Hadoop Pipes. First, one enables the user to run jobs with any executable. The second one is the Simplified Wrapper and Interface Generator SWIG tool, which helps to connect the C++ library with scripting languages.

2) *Tcpdump*: It is compatible with all most UNIX operating systems like LINUX. Tcpdump utilizes to enroll in network traffic. Tcpdump save the network traffic in diverse formats that help capture the packets using the specific criteria from the users. To analyses the network traffic Tcpdump running under the command line. Tcpdump users are permitted to display TCP/IP packets that are sent or received over the host network [18]. Libpcap library is running with Tcpdump to catch packages. Libpcap can read the packet file from the network interface, as well as it is displayed and intercept communication between computer devices or between users.

3) *CoralReef*: It is an overall package utilized for various programming languages, device drivers, classes, applications, and libraries. Generally, the coral reef applications are classified into two types: the first one begins with CoralReef and connect with raw packet data, the second one is work on aggregated flow data [19]. The coral reef was utilized for:

- Follow the internet traffic, which records in the Coral Reef trace file.
- It is helpful to give information about the hardware devices and their connection of trace files.
- It helps protect privacy by encoding the IP address.
- With the IP protocol and length of it, CoralReef hits reports packet and byte counts, specify the port of the TCP and UDP.
- Utilized with many programming languages.

4) *Snort*: Snort is one of the open network tools, which is a free tool developed and designed by Martin Roesch in 1998 [18]. It is utilized for intrusion detection systems and intrusion prevention. Snort used real-time traffic analysis and packet logging on Internet Protocol (IP) networks. It is also can search for content and match the content. Snort has three modes can configure by the Packet sniffer, recording the network traffic for better debugging, and intrusion handling system [20]. With the first mode sniffer, can read the network packets. The second mode, record the packets in the hard disk HDD. Lastly, the network intrusion mode is utilized to monitor and analyze internet traffic.

5) *Tstat*: It's one of an open-source tool that is utilized for traffic analysis. Tstat is an automated tool for passive monitoring, which developed to be a more complex tool to provide rich statistics and functionality [21]. Unlike other tools, it does not target the classification as the main goal. Tstat software tool is utilized for a wide analysis of Internet traffic at the flow level. Therefore, it is supporting lesser application protocols. Based on the last studies, Tstat act as a source of ground truth [22].

6) *L7 Filter*: Its open-source project, which was released in 2003, usually proved the classifier for Linux's Netfilter subsystem. It can classify the packages based on application layer data .generally L7 Filter has two versions: kernel module for Linux, and runs as a userspace program [23]. L7 Filter utilized for:

- Using the predefined signatures, L7 Filter can identify the applications.
- Has signatures of many applications such as TCP.
- Working compatible with traffic classification community.

7) *NetFlow*: To collecting the IP information of the traffic and monitoring network, Cisco has been developed a NetFlow network protocol. When the routers are utilized, the NetFlow can generate NetFlow records. NetFlow collector utilized to group the records data, which are exported from the router, then process it to perform the traffic analysis and presentation in simple format[24].NetFlow has the switching feature, which helps to export it UDP or SCTP protocols. NetFlow utilized for:

- Provide aggregation features using the NetFlow collector.
- It minimizes the cost and maximizes the performance.
- It's detecting the unuseful traffic and the analysis of network applications.
- Reduce the risk of performance.

G. Simulation and Configuration

The simulation was conducted by using five personal computers; one as a server and four virtual computers as network nodes. Each computer has 2GB memory and 100GB storage. Five types of data segmentation are 10 MB, 40MB, 64MB, 200MB, and500MB. The speed of the network is calculating in a megabit per second (Mbs) through based upon the network speed on the number of allocated PCs (100 Mbs/4).

The purpose of the proposed framework is to understand the effects of scalable options of Internet traffic measurement and analysis on the data transfer speed. Hadoop/ MapReduce can handle traffic measurement and analysis package according to reasons like possibility reprogrammed various programming, for example, C ++ and Java dialects, and similarity when utilized working frameworks control. Hadoop/ Map Reduce as well as the possibility of measuring features of the passing update, as the level of use and the provision of different people the scope and data network, and ability to be calculated based on other transfer protocols like TCP and UDP [6].

This study recommends the Hadoop/Map Reduce is chosen as an apparatus with the end goal of Internet movement estimation and examination. Hadoop /Map Reduce can reinforce the versatile estimation and examination frames due to its versatility of redesign the segments of the estimation techniques to bolster the adaptable to quantify and investigation operations from its adaptability of upgrade the lineaments of measure procedures. For example, arrange the system used to numerous classes and deliver compelling information divisions number rely on upon the information size and the system use.

Furthermore, the Hadoop/Map Reduce provides flexible traffic analysis while transfer the different transfer data through protocols like TCP with UDP. However, many weaknesses are still not covered well, such as how to classify the network usage low, medium, or heavy usage. What is the best segmentation number of data based on the networks usage level and data characteristics? Another issue is about the performance level transfer time of transfer protocols based on data type in this tool that needs to be addressed [25]. The main implication of this paper is applying adaptive Internet traffic measurement and analysis options based on the dynamic changes in network traffic environments. Therefore, the best options from scalable measurement and analysis options could be selected adaptively to assure the speed of data transfer. To clarify this implication, consider the following two scenarios.

Scenario # A: user X transfer documents of data size 500 MB to user Y through the network that is classified as a low traffic network. According to Hadoop/MapReduce options, this document could be transferred using a 64 MB packet size based on TCP protocol. The estimated transfer time is based on these options, and according to our experimental results, it is 11.1 seconds.

Scenario # B: user X transfer documents of data size 500 MB to user Y through a network that is classified as a low traffic network. According to our approach of scalable transfer options using Hadoop/Map Reduce, this document could be transferred using 200 MB packet size based on TCP protocol. The estimated transfer time is based on these options and, according to our experimental results, is 8.9 seconds. Based on the scenarios, it is clear the data should be transferred adaptively based on the most suitable decisions (transfer options) of Internet traffic measurement and analysis.

IV. CONCLUSION

This paper examines different methodologies of Internet movement estimation and examination. The focal points and hindrances of every instrument are dissected because of the execution adaptability of those devices based on element systems situations. Hadoop/Map-Reduce approach record upper hands over different apparatuses. Hadoop/Map Reduce can bolster the adaptable estimation and examination forms because of its adaptability of overhaul the components of the estimation procedures and activity investigation while exchanging the information.

Test more scalable options of Internet traffic measurement and analysis. Many scalable traffic measurement options could be tested to classify the network usage as five classes (for example: low, low, medium, substantial, and overwhelming), and provide a more scalable segmentation number of data packets. On the other hand, then many scalable traffic analysis options could be tested, such as transfer the data using the integration of transfer protocols (i.e., the integration of UDP and TCP may provide more effective results). Explore the testing based on various real environments: the transferred data nature is differing according to network users. For example, the users for educational purposes may transfer text data, while the users for social objectives may transfer videos and images. This requires careful analysis of the data sizes, types, and formats that are transferred for many purposes.

Based on our research Hadoop/MapReduce represents an appropriate tool due to many reasons like it's easy to write by various programming languages like JAVA and C++. Hadoop/MapReduce is compatible with many operating systems. It can handle the traffic measurement and analysis so on. Besides that, it's enhancing the possibility of updating the traffic measurement features like increase the coverage of the network use and providing more segmentation of the data. Moreover, by using Hadoop/MapReduce, the traffic analysis can perfect dynamically before transporting the data using various transporting protocols like TCP (Transmission Control Protocol) and UDP (User Datagram Protocol). Finally, based on our work, we recommend the Hadoop/MapReduce tool to utilize for internet traffic measurement and analysis. Mainly due to the flexibility of updating the features in the measurement process, Hadoop/MapReduce assists the scalable measurement and analysis.

REFERENCES

- [1] J. Ni, W. Weng, J. Chen, and K. Lei, "Internet Traffic Analysis Using Community Detection and Apache Spark," in *Proceedings - 2017 Int. Conf. on Cyber-Enabled Distributed Computing and Knowledge Discovery, CyberC 2017*, 2017, vol. 2018-Janua, pp. 213–219, doi: 10.1109/CyberC.2017.85.
- [2] E. Nazari, M. H. Shahriari, and H. Tabesh, "Big Data Analysis in Healthcare: Apache Hadoop, Apache spark and Apache Flink," *Front. Heal. Informatics*, vol. 8, no. 1, p. 14, 2019, doi: 10.30699/fhi.v8i1.180.
- [3] R. Y. Zhong, S. T. Newman, G. Q. Huang, and S. Lan, "Big Data for supply chain management in the service and manufacturing sectors: Challenges, opportunities, and future perspectives," *Comput. Ind. Eng.*, vol. 101, pp. 572–591, 2016, doi: 10.1016/j.cie.2016.07.013.
- [4] L. T. Ibrahim, R. Hassan, K. Ahmad, A. N. Asat, and H. Omar, "Online traffic measurement and analysis in big data: Comparative research review," *Am. J. Appl. Sci.*, vol. 13, no. 4, pp. 420–431, 2016, doi: 10.3844/ajassp.2016.420.431.
- [5] D. W. Jacob, M. F. M. Fudzee, M. A. Salamat, "A Conceptual Study on Generic End-Users Adoption of e-Government Services," *Int. Journal on Advanced Science, Engineering and Information Technology*, 7 (3), pp.1000-1006, 2017.
- [6] L. T. Ibrahim, R. Hassan, K. Ahmad, and A. N. Asat, "A study on improvement of internet traffic measurement and analysis using Hadoop system," in *2015 International Conference on Electrical Engineering and Informatics (ICEEI)*, 2015, pp. 462–466, doi: 10.1109/ICEEI.2015.7352545.
- [7] H. Lu, C. Hai-Shan, and H. Ting-Ting, "Research on Hadoop Cloud Computing Model and its Applications," *2012 Third Int. Conf. Netw. Distrib. Comput.*, pp. 59–63, 2012, doi: 10.1109/ICNDC.2012.22.
- [8] H. Cui, Y. Yao, K. Zhang, F. Sun, and Y. Liu, "Network Traffic Prediction Based on Hadoop," in *2014 International Symposium on Wireless Personal Multimedia Communications (WPMC)*, 2014, no. c, pp. 29–33, doi: 10.1109/WPMC.2014.7014785.
- [9] F. Dongyu, L. Zhu, and Z. Lei, "Review of hadoop performance optimization," in *2016 2nd IEEE International Conference on Computer and Communications (ICCC)*, 2017, pp. 65–68, doi: 10.1109/CompComm.2016.7924666.
- [10] K. Jeřábek and O. Ryšavý, "Big data network flow processing using apache spark," in *ACM International Conference Proceeding Series*, 2019, pp. 1–9, doi: 10.1145/3352700.3352709.
- [11] Y. Liu, H. Lu, X. Li, and D. Zhao, "An approach for service function chain reconfiguration in network function virtualization architectures," *IEEE Access*, vol. 7, pp. 147224–147237, 2019, doi: 10.1109/ACCESS.2019.2946648.
- [12] J. Abawajy, M. F. M. Fudzee, M. M. Hassan, M. Alrubalan, "Service Level Agreement Management Framework for Utility-oriented Computing Platform" *Journal of Computing*, 71 (11), pp. 4287–4303, 2015.
- [13] R. Pries, F. Wamser, D. Staehle, K. Heck, and P. Tran-Gia, "Traffic Measurement and Analysis of a Broadband Wireless Internet Access," in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, 2009, pp. 1–5, doi: 10.1109/VETECS.2009.5073890.
- [14] B. Xu and S. A. Kumar, "Big Data Analytics Framework for System Health Monitoring," in *Proceedings - 2015 IEEE International Congress on Big Data, BigData Congress 2015*, 2015, pp. 401–408, doi: 10.1109/BigDataCongress.2015.66.
- [15] Y. Lee & Y. Lee, "Toward scalable internet traffic measurement and analysis with Hadoop," *ACM SIGCOMM Comput. Commun. Rev.*, vol. 43, no. 1, pp. 5–13, 2013, doi: 10.1145/2427036.2427038.
- [16] N. A. Liu & J. Feng Liu, "Monitoring and analyzing big traffic data of a large-scale cellular network with Hadoop," *IEEE Netw.*, vol. 28, no. 4, pp. 32–39, 2014, doi: 10.1109/MNET.2014.6863129.
- [17] D. Quick and K. K. R. Choo, "Digital droplets: Microsoft SkyDrive forensic data remnants," *Futur. Gener. Comput. Syst.*, vol. 29, no. 6, pp. 1378–1394, 2013, doi: 10.1016/j.future.2013.02.001.
- [18] J. Yang, L. Wang, A. Lesh, and B. Lockerbie, "Manipulating network traffic to evade stepping-stone intrusion detection," *Internet of Things*, vol. 3–4, pp. 34–45, 2018, doi: 10.1016/j.iot.2018.08.011.
- [19] L. U. Laboshin, A. A. Lukashin, and V. S. Zaborovsky, "The Big Data Approach to Collecting and Analyzing Traffic Data in Large Scale Networks," *Procedia Comput. Sci.*, vol. 103, no. October 2016, pp. 536–542, 2017, doi: 10.1016/j.procs.2017.01.048.
- [20] A. Ibitola and O. B. Longe, "Analysis and Evaluation of Network-Based Intrusion Detection and Prevention System in an Enterprise Network Using Snort Freeware," *African J. Comput. ICT*, vol. 6, no. 2, pp. 169–184, 2013.
- [21] A. Finamore, M. Mellia, M. Meo, M. M. Munafo & D. Rossi, "Experiences of Internet traffic monitoring with tstat," *IEEE Netw.*, vol. 25, no. 3, pp. 1–7, 2011.
- [22] D. Adami, C. Callegari, S. Giordano & M. Pagano, "Skype-Hunter: A real-time system for the detection and classification of Skype traffic," *Int. J. Commun. Syst.*, vol. 25, no. 3, pp. 386–403, 2012.
- [23] C. Shen and L. Huang, "On detection accuracy of L7-filter and OpenDPI," in *2012 Third International Conference on Networking and Distributed Computing*, 2012, pp. 119–123.
- [24] A. Bakhshandeh, "An efficient user identification approach based on Netflow analysis," in *2018 15th Int. ISC (Iranian Society of Cryptology) Conf. on Information Security and Cryptology (ISCISC)*, 2018, pp. 1–5.
- [25] G. Poppe, F., Van den Bosch, S., de La Vallée-Poussin, P., Van Hove, H., De Neve, H., & Petit, "Choosing the objectives for traffic engineering in IP backbone networks based on Quality-of-Service requirements," in *International Workshop on Quality of Future Internet Services*, 2000, pp. 129–140.