International Journal of Pure and Applied Mathematics
Volume 116 No. 8 2017, 399-403
ISSN: 1311-8080 (printed version); ISSN: 1314-3395 (on-line version)
url: http://www.ijpam.eu
Special Issue

ijpam.eu

REVIEW OF MONITORING WEB TRAFFIC

[1]S.Pothumani, [2]C.Anuradha,[3]K.Lakshmikanthredddy
[1,2]Asst Professor, Department of CSE,BIST,BIHER, Bharath University,
[3]UG Student, Department of CSE, BIST,BIHER,BharathUniversity,Chennai,TN,INDIA
[1]pothumani@gmail.com,[2]anuradha.ak23@gmail.com,[3]kranthimanu143@gmail.com

Abstract: The essentially web traffic estimation and examination is vital to maintain a strategic distance from numerous issues of information exchanging through online networks, for example, information losing and moderate information exchanging as the quantity of web clients increment quickly in this world web traffic is additionally increment in computer network traffic estimation is the procedure measuring the sum and sort of traffic on a specific network web traffic estimation and investigation are for the most part use to describe and investigation of network utilization and client conduct to investigation this traffic different instruments are accessible yet they don't perform well when the traffic information measure increment because of development in web clients and transmission capacity hungry applications the measure of web traffic information created so enormous it requires versatile

Keywords: Traffic monitoring, Hadoop, mapreduce, HDFS, net flow.

1. Introduction

Network monitoring and measurement have received extra significance in modern-day day complex community formerly ,network administrator reveal simplest a limited number of network device or computer whose variety variety much less than a hundred community bandwidth can be just 10 or a hundred Mbps (Megabit according to 2d) ; However[1-2], now directors should deal with now not most effective better pace stressed out network (extra than 10 Gbps (Gigabit in keeping with sec) and ATM (Asynchronous Transfer Mode) network) but additionally wireless networks. They need extra state-of-the-art network visitors monitoring and analysis tools with a purpose to maintain the community device balance and availability which includes to restore community problems on time or to avoid network failure, to make certain the network protection electricity, and to make excellent decisions for network planning. When a network failure happens, tracking retailers must hit upon[3-4], isolate, and accurate malfunctions inside the network and likely get better the failure. Commonly, the marketers ought to warn the directors to fix the troubles within a minute. With the stable network, the directors' jobs stay to screen continuously if there's a hazard from either inner or outdoor network. Moreover[5-6], they must frequently test the network overall performance if the community devices are overloaded.

Before a failure due to the overload, information about community usage may be used to make a community plan for quick-term and long-time period future development. In laptop networks, community visitors size is the procedure of measuring the amount and form of site visitors on a particular network. Network analysis might be measured by means of active approach and passive strategies. Active strategies are greater intrusive however are arguably more accurate. Passive techniques are of much less community overhead and therefore can run within the heritage for use to cause community management movements[7-8]. A drawback of energetic size is that it can disturb the community by using injecting synthetic probe site visitors into the network and the principle drawback of the usage of this passive size is that he assumed that he “owns” all networks in the network traffic measurement they are two challenge like 1) Flow facts computation time 2) Single node failure. To deal with these mission, I need to put in force the internet visitors size and evaluation using MapReduce programming version of Hadoop framework. Apache Hadoop is an open supply software frame work for storage and huge scale processing of Netflowdatasets[9-10].

This proposed will offer a brand new method to measuring and analysis Internet visitors which is based totally on MapReduce Paradigm. This new approach will try to enhance the computational time, more fault tolerance of system and will take care of or cope with Bigdata at some stage in internet visitors size and analysis.

2. Related work

This will deliver the concept of how distinct network site visitors gear is working on their environment. After that we can discuss and classify existing approaches for Internet site visitors measurement and evaluation with their blessings and dilemma. This chapter is written according to literature survey finished by way of us for stopping disclosure of sensitive association policies from unauthorized access[13-14].

Internet site visitors size and evaluation has lengthy been used to represent network utilization and person behaviors'. In Internet traffic measurement and analysis, go with the flow-primarily based site visitors monitoring techniques are widely deployed in the course of Internet Service Providers (ISPs), due to the fact the volume of processed facts is reduced and lots of handy glide statistics gear are to be had. Cisco NeTFlow-without problems monitors flows passing via routers and switches without observing every packet. Unfortunately its format is not open and it has been designed best for IPv4 community tracking. Wire shark is not an intrusion detection system. It will now not warn you when a person does atypical matters to your community that he/she is not allowed to do and Wire shark will now not manage matters on the network [11-12].

Tcpdump device has a few drawback like- this tool can handiest recognition on what it find, mean Tcpdump can record on most effective what it finds inside the packets. If IP cope with is forged in the packets, Tcpdump has no potential to file anything else . Pandora FMS If you want to preserve for your eyes on offerings, programs and verbal exchange, this tool is best for this reason however Pandora FMS get struggled when we want to control large community environment. Angry IP scanner scans IP address in its port and locating live hosts and providing you statistics approximately them. There aren't any obvious drawbacks to mention but Sometimes Angry IP Scanner cannot discover open ports and could recollect them as filtered. The most important reason of Network Miner is facts series for future analysis (forensic proof analysis) in preference to accumulating facts regarding the visitors at the network. Information are grouped through host instead of via packets or frames[15-16] .

3. Big data

Big information is a time period for big facts units, a massive quantity of information available in complicated established or no form shape. These large quantities of information are generated with the useful resource of social media and networks, medical units, cell gadgets, sensor generation an networks. The procedure of studies into big amounts of statistics to show hidden styles and mystery correlations named as massive records analytics. If the records which is past to the garage ability and that is past to the processing electricity that data is calling Big Data.

4. Apache hadoop framework

• Hadoop Common: It having utilities that assist the alternative hadoop module.
• Hadoop dispensed File System (HDFS): a allocated record-gadget that shops information on commodity machines, supplying very high combination bandwidth across the cluster.
• Hadoop YARN: A framework for assignment scheduling and cluster beneficial useful resource manipulate[17-18].
• HadoopMapReduce: a programming model for large scale information processing

5. Hadoop distributed file system

HDFS stands for "Hadoop Distributed File System" and HDFS is highly scalable record tool. HDFS enables parallel studying and processing facts. HDFS has a master/slave architecture. An HDFS cluster includes a single NameNode[19-20], a grasp server that manages the file device namespace and regulates get proper of access to to documents by clients. In addition, there are a number of DataNodes, usually one according to node within the cluster, which manipulate garage connected to the nodes that they run on. HDFS exposes a report device namespace and permits man or woman records to be saved in files. Internally, a record is break up into one or greater blocks and people blocks are stored in a hard and fast of DataNodes. The NameNode executes report machine namespace operations like beginning, last, and renaming documents and directories. It additionally determines the mapping of blocks to DataNodes. The DataNodes are responsible for serving read and write requests from the file device's customers. The DataNodes also carry out block creation, deletion, and replication upon guidance from the NameNode

6. Map reduce

Map/lessen is a special shape of this type of DAG (Direct acyclic Graph) that's applicable in a good sized variety of use instances and it's a programming paradigm for processing huge datasets in disbursed environment.It is prepared as a "map" characteristic which rework a piece of statistics into a few quantity of key/fee pairs. Each of those elements will then be looked after by using the usage of their key and reach to the same node, wherein a

"lessen" function is use to merge the values (of the identical key) into a unmarried end result.

7. Design of experiments

For the experiments we used, Wire shark model 1.12.Four ,Operating System: sixty 4-bit Ubuntu 14.04 LTS with Intel Core i5 CPU @ 2.30 GHz × four and 3 GB of RAM. Also we're took the dataset in .Pcp report Number of packets:305916,22189601

8. Experiment and result

Protocol Summarization analysis for pcap record in Wire shark tool. We feed that packets in one of the famous network evaluation device "Wireshark" and this device took 5.727 seconds to carry out protocol summarisation evaluation. Then we took every other pcap record, which consist of 22189601 packets in "wireshark" tool.

Table 1. Map Reduce Program

No of packets & size of file	**Protocol summarization analysis time in wire shark (sec)**
305916(250MB)	5.727
22189601(1GB)	111

9. Problem gap

From the protocol summarization take a look at, we will without trouble remember that whilst information is small in size, wireshark tools result is ideal[21], but as we can see that once facts is going big, the evaluation result is time ingesting, so we've got were given come up with some other approach for same trouble, that is Hadoop Technology. Hadoop is an open deliver framework given by using apache software program foundation for storing huge dataset and for processing huge dataset with the cluster of commodity hardware. So essentially Hadoop is for storing and processing bigdata.

10. Conclusion

Internet site visitors size and evaluation had been normally achieved on a high overall performance server that collects and take a look at packets. When we display a big amount of web site visitors information for detailed statistics, it isn't always clean to cope with Tera or Pera-bytes website online visitors with unmarried server. Scalable Internet visitors ameasurement and assessment is hard due to the truth a massive facts set require matching computing and garage useful aid. So we delivered Hadoop, an open-source computing platform of MapReduce and a dispensed report gadget, has become a well-known infrastructure for large information analytics because it permits scalable records processing and storage issuer on a disbursed computing device.

References

[1]. Abrahamsson, H., 2008. Internet traffic management.PhD.Thesis, MälardalenUniversity.Adami, D., C. Callegari, S. Giordano,
[2].M. Pagano and T. Pepe, 2012.
[3]. J. Commun. Syst., 25: 386-403. DOI: 10.1002/dac.1247 Ahuja, R.K., T.L. Magnanti and J.B. Orlin, 2015. Network Flows.
[4]. Edn.,BiblioLife, ISBN-10: 297491769, pp: 226. Lena T. Ibrahim et al.
[5]. American Journal of Applied Sciences 2016, 13 DOI: 10.3844/ajassp.2016.420.431 430 Alcock, S. and R. Nelson, 2013. Measuring the accuracy of open-source payload-based traffic classifiers using popular Internet applications. Proceedings of the 38th Conference on Local Computer Networks Workshops, Oct. 21-24, IEEE Xplore Press, Sydney, NSW, pp: 956-963. DOI: 10.1109/LCNW.2013.6758538 Angryip, 2014. Angry IP Scanner.http://angryip.org Arlitt, M.
[6] Udayakumar R., Kaliyamurthie K.P., Khanaa, Thooyamani K.P., Data mining a boon: Predictive system for university topper women in academia, World Applied Sciences Journal, v-29, i-14, pp-86-90, 2014.
[7] Kaliyamurthie K.P., Parameswari D., Udayakumar R., QOS aware privacy preserving location monitoring in wireless sensor network, Indian Journal of Science and Technology, v-6, i-SUPPL5, pp-4648-4652, 2013.
[8] BrinthaRajakumari S., Nalini C., An efficient cost model for data storage with horizontal layout in the cloud, Indian Journal of Science and Technology, v-7, i-, pp-45-46, 2014.
[9] BrinthaRajakumari S., Nalini C., An efficient data mining dataset preparation using aggregation in relational database, Indian Journal of Science and Technology, v-7, i-, pp-44-46, 2014.
[10] Khanna V., Mohanta K., Saravanan T., Recovery of link quality degradation in wireless mesh networks, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4837-4843, 2013.
[11] Khanaa V., Thooyamani K.P., Udayakumar R., A secure and efficient authentication system for distributed wireless sensor network, World Applied Sciences Journal, v-29, i-14, pp-304-308, 2014.

[12] Udayakumar R., Khanaa V., Saravanan T., Saritha G., Retinal image analysis using curvelet transform and multistructure elements morphology by reconstruction, Middle - East Journal of Scientific Research, v-16, i-12, pp-1781-1785, 2013.
[13] Khanaa V., Mohanta K., Saravanan. T., Performance analysis of FTTH using GEPON in direct and external modulation, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4848-4852, 2013.
[14] Kaliyamurthie K.P., Udayakumar R., Parameswari D., Mugunthan S.N., Highly secured online voting system over network, Indian Journal of Science and Technology, v-6, i-SUPPL.6, pp-4831-4836, 2013.
[15] Thooyamani K.P., Khanaa V., Udayakumar R., Efficiently measuring denial of service attacks using appropriate metrics, Middle - East Journal of Scientific Research, v-20, i-12, pp-2464-2470, 2014.
[16] R.Kalaiprasath, R.Elankavi, Dr.R.Udayakumar, Cloud Information Accountability (Cia) Framework Ensuring Accountability Of Data In Cloud And Security In End To End Process In Cloud Terminology, International Journal Of Civil Engineering And Technology (Ijciet) Volume 8, Issue 4, Pp. 376–385, April 2017.
[17] R.Elankavi, R.Kalaiprasath, Dr.R.Udayakumar, A fast clustering algorithm for high-dimensional data, International Journal Of Civil Engineering And Technology (Ijciet), Volume 8, Issue 5, Pp. 1220–1227, May 2017.
[18] R. Kalaiprasath, R. Elankavi and Dr. R. Udayakumar. Cloud. Security and Compliance - A Semantic Approach in End to End Security, International Journal Of Mechanical Engineering And Technology (Ijmet), Volume 8, Issue 5, pp-987-994, May 2017.
[19] Thooyamani K.P., Khanaa V., Udayakumar R., Virtual instrumentation based process of agriculture by automation, Middle - East Journal of Scientific Research, v-20, i-12, pp-2604-2612, 2014.
[20] Udayakumar R., Thooyamani K.P., Khanaa, Random projection based data perturbation using geometric transformation, World Applied Sciences Journal, v-29, i-14, pp-19-24, 2014.
[21] Udayakumar R., Thooyamani K.P., Khanaa, Deploying site-to-site VPN connectivity: MPLS VsIPSec, World Applied Sciences Journal, v-29, i-14, pp-6-10, 2014.