

Hadoop Based Big Data Traffic Handling in Ample Cellular Network

K.Aishwarya¹, Ms.B.M.Brinda², ³Ms.K.B.SriSathya

¹ PG Scholar, Dept. of Information Technology, Vivekanandha College of Engineering for Women Tiruchengode – 637205, India

^{2,3} Asst. Prof, Dept. of Information Technology, Vivekanandha College of Engineering for Women Tiruchengode – 637205, India

Abstract

Big Data creates major problems with large-volume of traffic, growing data access, complex data sets with numerous, autonomous sources. With the quick development of networking, data storage, and also the data assortment capability, big data is currently increasing rapidly in science and engineering domains, as well as physical, biological and biomedical sciences. The size of those data will become impractical. Hence, in such cases, the analyst should be capable of specializing within the informational data while ignoring the noise data. These difficulties complicate the multichannel data analysis when compared with the analysis of single-channel data. Traffic management in internet is difficult because a large data set requires matching computing and storage resources. The proposed work provides security for traffic data in ample cellular network and big data traffic handling based on hadoop. Due to handling of large data in efficient manner, the proposed work provides model process high traffic data with high performance.

Keywords – Big Data, Hadoop, Map Reduce, Traffic handling.

I. INTRODUCTION

The emergence of cellular networks provides large volume of data traffic carried by cellular networks because of the rapidly increasing subscriber population, capable user terminals in cellular communication and the expanding range of mobile applications. Understanding of Internet traffic in ample cellular networks is important for network design, troubleshooting, performance prediction, and optimization. As the number of network elements has become more difficult for Internet Service Providers (ISP) to gather and analyze with efficiency an outsized information set of raw packet dumps, flow records, activity logs, and security. To occupy demands for the deep analysis of ever-growing net traffic information, ISPs want a traffic management system wherever the computing and storage resources are often scaled out.

One massive cellular operator has reportable a growth of 8000% of cellular data traffic over the past four years and it's expected to grow to 15.8 exabytes per month by 2016. To deal with this explosive growth and best serve their customers, operators have been forced to have a higher understanding of the character of traffic carried by cellular networks. To handling maasive amount of traffic in cellular network the hadoop framework and MapReduce programming model can be proposed and provide the security for high traffic data.

II. EXISTING SYSTEM

The traffic acquisition devices collect traffic data at low and packet levels. These both need the network management system to employ an oversized storage system and superior computing server to store and analyze the traffic data from an enormous number of network devices. However, as networks grow exponentially; administrators face the big data challenges such as managing and process an enormous quantity of traffic data. The existing network measurement systems, that typically consider a centralized high-performance server, will not handle this large amount of data. Due to Big data traffic, possible for DDos attack which create a system or server out of stock for legitimate users and, finally, degrades the service. DDos attack slows down the performance of the cellular network while accessing the data.

III. RELATED WORK

Over the past few decades, plenty of tools are developed and wide used for Internet traffic monitoring and analysing. The emergence technologies MapReduce and GFS [1, 2] which are the hadoop programming model can be used currently. Apache Hadoop [3], which was sponsored by Yahoo!, it is to provide open-source distributed computing framework with MapReduce as the programming model and the Hadoop Distributed File System (HDFS) as its distributed file system. Commonly, Hadoop is used by Yahoo!, Facebook, IBM and Twitter to develop and

execute applications for huge data sets[4]. RIPE [5] provides a library based on Hadoop to analyze PCAP files. They made the first attempt to analyze the raw packet data in Internet. Authors in [6, 7] investigated how to analyze large volumes of network data by Hadoop-based tools like MapReduce and Pig. The most popular tool for capturing and analyzing packet traces with libpcap is Tcpdump [8]. DDoS attacks and defense methods [9] proposed to fight with them.

IV. PROPOSED SYSTEM

A. Cellular Network

The Cellular Network has been classified into two types: Core network and Access network. A core network is the central part of a cellular network which provides various services to customers who are connected by the access network. The proposed model is deployed in a core network. The wireless data set was collected on a Gn interface between a gateway GPRS Support Node (GGSN) and a Serving GPRS Support Node (SGSN) during a cellular network.

Cellular networks offer a number of desirable features:

- More capacity than a single large transmitter, since the same frequency can be used for multiple links as long as they are in different cells
- Mobile devices use less power than with a single transmitter or satellite since the cell towers are closer
- Larger coverage area than a single terrestrial transmitter, since additional cell towers can be added indefinitely and are not limited by the horizon.

B. Hadoop Ecosystem

Hadoop is an open-source framework which provides the distributed data processing that allows massive amount of system to work independently. By using hadoop the storage space and process power of traffic data can be increased. Hadoop has a master and multiple slave nodes. Hadoop Distributed File System (HDFS) is distributed and portable file system in Hadoop. HDFS is specifically designed for storing vast amount of data, so it is optimized for storing/accessing a relatively small number of very large files compared to traditional file systems where are optimized to handle large numbers of relatively small files. MapReduce is programming model which supports Hadoop.

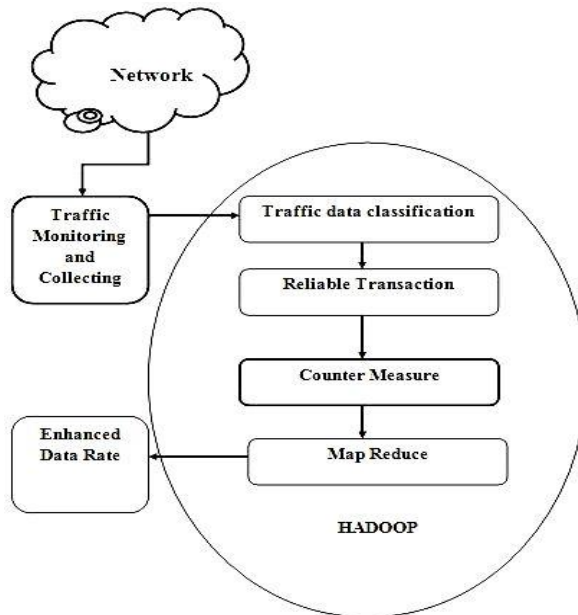


Fig: System Architecture

C. Traffic Monitoring and Collecting

The process of collecting and monitoring a big traffic data in cellular network is a difficult process. The constant smooth operation in the network is necessary for efficient performance.

The process in traffic monitoring and collecting:

- *Flow Collector* – responsible for flow filtering and collection of flow information history which switched within an interface.
- *Flow Caching* – it collect an IP data flow that enters on interface and prepares the data for exportation.
- *Data Analyzer* – responsible for data presentation where data collected can be used for various purposes other than monitoring network traffic.

D. Traffic Data Classification

The monitored and collected traffic data can be classified using C4.5 which can run parallel in mutiple system.It gives data structure customized for distributed parallel computing environment. C4.5 classifier conducts recursive partitioning over instance space.

E. Trouble shooting

Trouble shooting in cellular network using hadoop can be done by ATPG (Automatic Test Packet Generation). It sends the data from source to destination and checks whether there is any misbehaving node in the network which cause on attack in the cellular network. ATPG can detect the functional and

performance problems. ATPG periodically sends a set of test packets. If test packets fail, ATPG pinpoints the fault(s) that cause the problem.

F. Traffic Rate Enhancement

Due to the big data traffic rate can be unavailable for legitimate user in cellular network. For the enhancement of traffic rate the Map reduce and Multi party access control technique can be used. Based on the amount of user usage of data the traffic rate gets increase. If suppose more number of user enters the network due to roaming or mobility of node the usage of data rate will increase automatically it leads to increase the speed of the network This increases the availability of data to all the legitimate user in the network.

V. CONCLUSION AND FUTURE ENHANCEMENT

To occupy demands for the deep analysis of ever-growing net traffic information, ISPs want a traffic management system wherever the computing and storage resources are often scaled out. To handling maasive amount of traffic in cellular network the hadoop framework and MapReduce programming model can be proposed and provide the security for high traffic data. Based on this paper, algorithms have designed and implementes MapReduce programs for network traffic handling and provide the security for traffic data.

REFERENCES

- [1] S. Ghemawat, H. Gobioff, and S. T. Leung, "The Google File System," ACM SIGOPS Operating Systems Rev., vol. 37, no. 5, 2003.
- [2] J. Dean and S. Ghemawat, "MapReduce: Simplified Data Processing on Large Clusters," Commun. ACM, vol. 51, no. 1, 2008, pp. 107–13.
- [3] Hadoop, <http://hadoop.apache.org/>.
- [4] T. White, Hadoop: the Definitive Guide, O'Reilly, 3rd ed., 2012.
- [5] Large-Scale PCAP Data Analysis Using Apache Hadoop, <https://labs.ripe.net/Members/wnagele/large-scale-pcapdata-analysis-using-apache-hadoop>, accessed 9 Nov., 2013.
- [6] T. Samak, D. Gunter, and V. Hendrix, "Scalable Analysis of Network Measurements with Hadoop and Pig," IEEE Network Operations and Management Symp., 2012, pp. 1254–59.
- [7] T. P. D. B. Vieira, S. F. D. L. Fernandes, V. C. Garcia, "Evaluating MapReduce for profiling Application Traffic," Proc. 1st ACM Wksp. High Performance and Programmable Networking, 2013, pp. 45–52.
- [8] Tcpdump, <http://www.tcpdump.org>.
- [9] B. B. Gupta, J. Joshi, R. C. Misra, and Manoj (2009), 'Defending against Distributed Denial of Service Attacks: Issues and Challenges', Information Security Journal: A Global Perspective, 18: 5, 224 — 247.
- [10] J. T. Morken, "Distributed network processing using the map-reduce model," Ph.D. dissertation, Norwegian University of Science and Technology, 2010.
- [11] Y. Lee and Y. Lee, "Toward Scalable Internet Traffic Measurement and Analysis with Hadoop," ACM SIGCOMM Comp. Commun. Rev., vol. 43, no. 1, 2012, pp. 5–13.
- [12] Large-Scale PCAP Data Analysis Using Apache Hadoop, <https://labs.ripe.net/Members/wnagele/large-scale-pcapdata-analysis-using-apache-hadoop>, accessed 9 Nov., 2013.
- [13] T. P. D. B. Vieira, S. F. D. L. Fernandes, V. C. Garcia, "Evaluating MapReduce for profiling Application Traffic," Proc. 1st ACM Wksp. High Performance and Programmable Networking, 2013, pp. 45–52.
- [14] F. Ricca et al., "An Empirical Study on Keyword-Based Web Site Clustering," Proc. 12th IEEE Int'l. Wksp. Program Comprehension, 2004, pp. 204–13.
- [15] The R Project for Statistical Computing, <http://www.r-project.org/>, accessed 9 Nov., 2013.
- [16] J. Liu and N. Ansari, "Identifying Website Communities in Mobile Internet Based on Affinity Measurement," Computer Commun., vol. 41, 2014, pp. 22–20.
- [17] G. Chittaranjan, J. Blom, and D. Gatica-Perez, "Who's Who with Big- Five: Analyzing and Classifying Personality Traits with Smartphones," IEEE 15th Annual Int'l. Symp. Wearable Computers, 2011, pp. 29–36.
- [18] D. Tanasa D and B. Trousse, "Advanced Data Preprocessing for Intersites Web Usage Mining," IEEE Intelligent Systems, vol. 19, no. 2, 2004, pp. 59–65.
- [19] Wireless Universal Resource FiLe Open Source Project (WURFL), <http://wurfl.sourceforge.net/>, [Online; accessed 9-Nov.-2013].