Online Internet Traffic Monitoring System Using Spark Streaming

ABSTRACT

- Owing to the explosive growth of Internet traffic, network operators must be able to monitor the entire network situation and efficiently manage their network resources.
- Traditional network analysis methods that usually work on a single machine are no longer suitable for huge traffic data owing to their poor processing ability.
- Big data frameworks, such as Hadoop and Spark, can handle such analysis jobs even for a large amount of network traffic.
- However, Hadoop and Spark are inherently designed for offline data analysis. To cope with streaming data, various stream-processing-based frameworks have been proposed,
- such as Storm, Flink, and Spark Streaming. In this study, we propose an online Internet traffic monitoring system based on Spark Streaming.
- The system comprises three parts, namely, the collector, messaging system, and stream processor. We considered the TCP performance monitoring as a special use case of showing how network monitoring can be performed with our proposed system.
- We conducted typical experiments with a cluster in standalone mode, which showed that our system performs well for large Internet traffic measurement and monitoring.

EXISTING SYSTEM

- Traditional network analysis methods that usually work on a single machine are no longer suitable for huge traffic data owing to their poor processing ability.
- Big data frameworks, such as Hadoop and Spark, can handle such analysis jobs even for a large amount of network traffic.
- However, Hadoop and Spark are inherently designed for offline data analysis. To cope with streaming data, various stream-processing-based frameworks have been proposed, such as Storm, Flink, and Spark Streaming.

DISADVANTAGE

- Hadoop and Spark are inherently designed for offline data analysis.
- this is difficult nowadays due to the huge scalability of networks and the huge amount of traffic to be analyzed.
- Existing approaches take advantage of big data frameworks to improve processing efficiency.
- However, these approaches mainly focus on offline data analysis.

PROPOSED

- we propose an online Internet traffic monitoring system based on Spark Streaming, which is a big data platform that can efficiently process a huge amount of traffic data so that we can monitor the network status in real time and is robust enough so as to suffer a failure without aborting the entire monitoring process Our contributions in this study are as follows:
- We propose a distributed architecture as an online Internet traffic measurement and monitoring system.
- We implement a parallel algorithm for monitoring TCP performance parameters, such as delay and retransmission ratio with a very short delay.
- We conduct typical experiments showing that the proposed system is feasible and efficient

ADVANTAGE

- We conducted typical experiments with a cluster in standalone mode, which showed that our system performs well for large Internet traffic measurement and monitoring.
- we proposed an online Internet traffic monitoring system that utilizes Spark Streaming. Extensive experimental results show that our system achieved good performance and robustness.

HARDWARE REQUIREMENTS

- Processor – Pentium –III
- Speed – 1.1 Ghz
- RAM – 256 MB(min)
- Hard Disk – 20 GB
- Floppy Drive – 1.44 MB
- Key Board – Standard Windows Keyboard
- Mouse – Two or Three Button Mouse
- Monitor – SVGA

SOFTWARE REQUIREMENTS

- Operating System : Windows 8
- Front End : Java /DOTNET
- Database : Mysql/HEIDISQL

CONCLUSION

- With the growth of Internet traffic, traditional network analysis methods that work on single machines are no longer suitable.
- Existing approaches take advantage of big data frameworks to improve processing efficiency.
- However, these approaches mainly focus on offline data analysis. In this study, we proposed an online Internet traffic monitoring system that utilizes Spark Streaming.
- We demonstrated that Internet measurement and monitoring can be treated as a stream analysis problem and can be handled via a streaming processing platform.
- Extensive experimental results show that our system achieved good performance and robustness.

FUTURE WORK

- In future, we will implement collectors to capture packets from switches through port mirroring so that our system can analyze all the traffics passing through monitored networks.
- Finally, we will test its performance in practice and compare it with some traditional single server systems in terms of scalability and reliability.

REFERENCE

[1] Cisco Visual Networking Index, Forecast and methodology, 2016-2021, White Paper, San Jose, CA, USA: Cisco, 2016.

[2] Y. Lee, W. Kang, and H. Son, An Internet traffic analysis method with MapReduce, inProc. 2010 IEEE/IFIP Network Operations and Management Symposium Workshops (NOMS Wksps), Osaka, Japan, 2010, pp. 357–361.

[3] D. Brauckhoff, B. Tellenbach, A. Wagner, M. May, and A. Lakhina, Impact of packet sampling on anomaly detection metrics, inProc.6th ACM SIGCOMM Conf.Int. Measurement, Rio de Janeriro, Brazil, 2006, pp. 159–164.

[4] Y. Y. Qiao, Z. M. Lei, L. Yuan, and M. J. Guo, Offline traffic analysis system based on Hadoop, J. China Univ. Posts Telecommun., vol. 20, no. 5, pp. 97–103, 2013.

[5] Hadoop, http://hadoop.apache.org/, 2017