Scalable Flow Monitoring for Data Center Network

A Project Report Submitted in the partial fulfillment of the
requirements for the award of degree of

Master of Technology
in
Computer Science

By
Nirmoy Das

Department of Computer and Information Sciences
University of Hyderabad
Hyderabad, India

June, 2008

CERTIFICATE

This is to certify that the project work entitled "**Scalable Flow Monitoring for Data Center Network**" being submitted to University of Hyderabad by **Nirmoy Das** (Reg. No. 11MCMT20), in partial fulfillment for the award of the degree of Master of Technology in Computer Science, is a bona fide work carried out by her under my supervision.

Ms. Anupama Potluri
Project Supervisor,
School of CIS,
University of Hyderabad

Dean ,
School of CIS,
University of Hyderabad

To,

My Parents and Supervisor.

Contents

Chapter 1

Related Work

Flow monitoring protocol like netflow and sflow can provide important information about the traffic that passes through a network. How ever contemporary computer networking is out-spacing out ability to monitor them efficiently. As data centers are getting virtualized with virtual software switches and scaling to thousands of node it is our immediate requirement to have monitoring to that scale efficiently. There are few solution that tries to provide some methods to have scalable flow monitoring in data centers.

1.1   EMC2[1]

EMC2 is flow collector that stay in side virtual switches in the hosts. Monitoring at server virtual switch is scalability due to its distributed nature.

1.1.1   Architecture

EMC2 is a multi-threaded application that spawns parser thread for accepting sFlow/NetFlow packet. EMC2 maintains a 2-level in-memory hash table that contains flow records. Flow-ID is primary key for in-memory hash table that maps another hash table where timestamp is the key and flow record is value. Flow record contains number of packets, number of bytes and optional path vector.

1.1.2   Deduplication

Deduplication avoids adding of same flow in the flow table reported by multiple Vswitch for the same flow. EMC2 uses simple heuristics to to detect duplicate flow.

Algorithm 1: Detect duplicate

---

    if $flow - ID$ not exist **then**

      add flow to the flow table.

      return

    else

      if Same exporter **then**

        update the flow table

        return

      else

        report duplicate flow

        update path vector

        return

      end if

    end if

1.1.3   Data Rate Prediction in Presence of Sampling

EMC2 predict data rate my multiplying length of the packet with sampling rate given in flow packet. It can also report low sampling rate by accumulating samples from different exported devices.

Bibliography

[1] S. B. Vijay Mann, Anilkumar Vishnoi, "Living on the edge: Monitoring network flows at the edge in cloud data center," in *Communication Systems and Networks (COMSNETS), 2013 Fifth International Conference.*

[2] Y. L. Yeonhee Lee, "Toward scalable internet traffic measurement and analysis with hadoop," *ACM SIGCOMM Computer Communication Review (CCR)*, vol. 43, pp. 5–13, jan 2013.